

Pharmacogenetics to come

Genetically selected medicine has been much hyped but has significant potential. Regulation and treatment will depend on pharmaceutical companies more readily sharing genetic data.

We all know someone who will laugh immoderately at bad jokes after a small glass of champagne, and someone else who can down several pints of beer and yet still discuss celestial mechanics with intellectual clarity. This is happy-hour pharmacogenetics in action.

Clinical pharmacogenetics works on the same principle — that people react individually to different drugs. It is a less light-hearted affair. Every year, hundreds of thousands of deaths worldwide can probably be attributed to the side effects of drugs; and hundreds of thousands of patients will take drugs that for them have no effect at all.

Thanks to new genomic technologies, it is getting easier and cheaper to identify the subtle genetic differences — ‘SNPs’ (single nucleotide polymorphisms) — that are responsible for such diverse responses (see page 760). Drug-regulatory agencies are now confronted with the problem of how to respond, and the challenge is not an easy one.

The popular, and much-hyped, image of a straightforward glide into perfect, personalized medicine is way off the mark. No one yet knows how predictive SNPs will be in identifying individuals who are likely to suffer side effects, or who may not benefit, from a particular drug. In some cases, when a polymorphism in a single gene dictates potentially lethal side effects in a drug for a serious disease, the benefits clearly outweigh the costs. But in most cases, large numbers of SNPs are likely to be required to profile those who may benefit or suffer — pushing up the cost without necessarily delivering a level of predictive certainty that will put physicians at ease. In treatment of non-life-threatening diseases, who decides when it is no longer worth incurring costs just to derive a slightly increased certainty of efficacy?

Regulatory agencies are now trying to decide what genomic data they should demand from pharmaceutical companies wanting to bring new drugs onto the market. In mid-November, the opinion-setting US Food and Drug Administration (FDA) will hold its second

workshop on pharmacogenetics in drug-regulatory decision-making. It aims to define guidelines to encourage drug companies to submit genomic data, reassuring them that FDA staff will not try to make predictions of possible toxicity from raw data that few are as yet used to interpreting. The agency's first meeting, in May last year, exposed the conservatism of many big drug firms. They are afraid of being forced to abandon their traditional blockbuster approach of drug development in favour of the potentially less lucrative method of targeting therapy to restricted groups of patients most likely to benefit. But they are enticed by the possibility of greatly reduced costs for clinical trials.

Conscious of the industry's fears, the FDA wants to avoid disincentives to producing new drugs. And nobody knows how the data will be used. Hence the non-threatening and vague terminology — ‘guidelines’, ‘encouragement’, ‘reassurance’.

But openness will have to increase. We are starting to learn more about the predictive value of SNPs from academic clinical studies on approved drugs. This information will slowly accrue. As the price of identifying SNPs falls yet further, physicians should expect to receive more pharmacogenetic data from drug companies, along with appropriate genetic tests, to help them select the best therapy for an individual patient. The smartest physicians will use this information appropriately — while not forgetting that a full and accurate description of clinical symptoms and exact diagnosis are equally important in interpreting such data and selecting therapy. Weakness in clinical data collection is universally acknowledged as a serious hindrance to pharmacogenetics.

The pharmacogenetic learning curve is not particularly steep — experience of the principles has been gathering slowly over the past half-century — but it will be long and grinding for all sides. The stakes are high and the future uncertain. There is no doubt that pharmacogenetics will improve therapeutics, but it will arrive gradually, and will not provide a panacea. ■

New access for agriculture

A United Nations scheme launched last week extends unrestricted access to *Nature's* content within developing countries.

It is gratifying, as staff at *Nature* have found, to meet researchers in infectious diseases in the world's poorer nations and discover that they are making use of their privileged access to our content. Since March 2002, researchers, policy-makers, educators and others in more than 1,000 institutions in 100 or so developing countries have been receiving *Nature* free of charge online as part of the Health InterNetwork Access to Research Initiative (HINARI).

Led by the World Health Organization, this United Nations initiative includes all of the journals published by Nature Publishing Group (NPG), and 2,000 or more from other publishers (for details see www.healthinternetwork.org/src/eligibility.php). Web statistics show that significant and increasing use is being made of this accessibility, for which the HINARI secretariat acts as gatekeeper. It includes all countries whose annual GNP per head is less than US\$3,000. (Publishers may charge reduced prices for access in the less impoverished of these countries, but NPG has elected to provide free access for all.)

An equivalent scheme for researchers in agriculture was launched on 14 October by the Food and Agriculture Organization of the United Nations. Inevitably, it has an acronym: AGORA (Access to Global Online Research in Agriculture) — see www.aginternetwork.org/en/about.php. As with HINARI in its first phase, all publishers sign up to providing free access to researchers and others in countries where the GNP is less than \$1,000 per capita (there are 69 eligible countries — see www.who.int/library/reference/temp/eligible_countries.pdf).

It is worth adding that *Nature* actively supports another free-access distributor of its content that is targeting the developing world: SciDev.Net (see www.SciDev.Net), financed by foundations and government departments for international development.

We at *Nature* are delighted to contribute to the principal goal of AGORA: to increase the quality and effectiveness of agricultural research and training in low-income countries, and thereby to improve food security. ■





Steroid shock
Standard test fails to
spot performance-
enhancing drug
p752



On the run
Atlantic salmon
under threat from
farm-bred cousins
p753



Tactical move
Standards lab faces
up to possible
military control
p754



Sweet relief
Supreme court rules
in favour of medical
marijuana use
p756

Biosafety trials darken outlook for transgenic crops in Europe

Jim Giles, London

The results are in from the largest investigation so far into the ecological impact of transgenic crops — and they're bad news for agricultural biotechnology.

Britain's Farm Scale Evaluations, published on 16 October, show that two genetically modified crops — spring oilseed rape and beet — are likely to have harmful impacts on farmland biodiversity. Researchers say the levels of weeds, seeds and insects in fields of transgenic crops were lower than those in plots of conventional varieties, and that this could have a knock-on effect on the birds and small animals that feed off these populations.

Although the problems are caused by the herbicide-spraying regime associated with the crops, rather than the crops themselves, the results are likely to make it politically impossible for the British government to license transgenic crops in the immediate future, many observers say. That's a blow for supporters of the technology in the United States, who had been looking to Britain for potential support in their attempts to persuade Europe to accept the technology.

The trials, which took place between 2000 and 2002 and are published as eight papers in the *Philosophical Transactions of the Royal Society B*, compared conventional and transgenic varieties across 200 plots.

Positive results for a third crop — maize (corn) — have been called into doubt, as the weed-killer used on most of the conventional plants is to be phased out. But the other results have not been directly challenged by most supporters of the technology. The data show that the number of seeds on the ground in the plots of transgenic oilseed rape and beet was one-third to one-sixth lower than in the conventional plots. Levels of some insects and weeds were also lower. "We could see a long-term decline in weeds that feed birds," says Les Firbank, a land-use specialist at Lancaster University, who led the trials.

Firbank and the other authors stress that it is the herbicide-spraying regime, not the genetic modification, that is the root of the problem. Herbicide-resistant crops are engineered to resist broad-spectrum weed-killers that remove almost all weeds from a field.



Scaled down: a study finds reduced levels of insects and seeds on which birds and small animals feed.

During the farm-scale evaluations, farmers sprayed the crops once or twice with a broad-spectrum herbicide. This reduces the labour required for conventional weed management, which involves repeatedly applying less powerful weed-killers. But the more powerful herbicide used with the transgenic crops also removes more weeds, as well as the seeds they produce.

Representatives of the agriculture industry point out that this leaves open the possibility that another herbicide-spraying regime might have lessened the impact on biodiversity while still reducing farmers' labour. "This was a test of management systems," says Craig Stevenson, head of government affairs at the London office of the agricultural biotechnology company Monsanto. "These can be changed very easily. We're still confident that transgenic crops can bring benefits."

But given the intense public opposition to transgenic agriculture, the chances of commercializing herbicide-resistant crops in the short term are slim (see *Nature* 425, 656–657; 2003). Many surveys of public attitudes,

including a government-commissioned public debate, have recorded opposition to commercialization, and the field-trial results are likely to strengthen this sentiment.

No decision will be made until a panel of scientific advisers has considered the results for the government. But environment minister Elliot Morley has already said that no commercial planting will take place in 2004.

In the meantime, work on better spraying regimes continues. Alan Dewar, an entomologist at Broom's Barn Research Station in Bury St Edmunds, Suffolk, and a researcher on the farm-scale evaluations, has studied the effect of delaying spraying beet until later in the crop's life, and of retaining herbicide-free strips between rows of crops. The experiments, which are partly funded by Monsanto, found that the effects on biodiversity may be no worse than with conventional crops. "But when I saw the headlines last week I wondered if we would ever get the chance to prove it," he says.

Additional reporting by Michael Hopkin.
♦ www.defra.gov.uk

Drugs bust reveals athletes' secret steroid

Jonathan Knight, San Francisco

Underground chemists may be concocting performance-enhancing drugs faster than drug testers can devise ways to spot them. That is the fear of sports physicians after revelations that dozens of elite athletes have been using an undetectable anabolic steroid to build muscle.

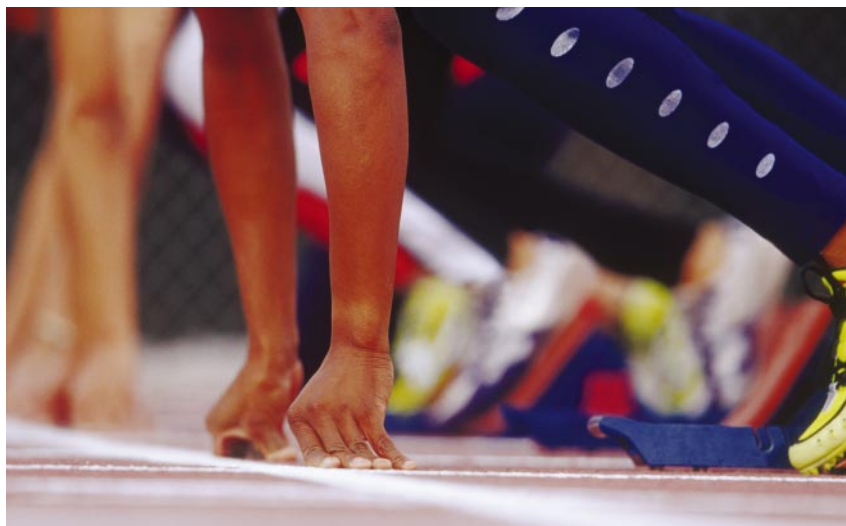
The US Anti-Doping Agency (USADA) received an anonymous tip-off in June that athletes were taking an undetectable steroid. The informer provided a used syringe of the substance, which the agency sent to the International Olympic Committee's drug-testing lab at the University of California, Los Angeles (UCLA).

The lab identified the compound as tetrahydrogestrinone (THG) a close relative of the anabolic steroid gestrinone, which is banned by the Olympic committee. Urine samples from some 20 athletes — who have not yet been publicly identified — have since tested positive.

This case is alarming for the authorities because the compound evades the standard two-step urine test for steroids. In this test, gas chromatography isolates steroids and a few other compounds, and mass spectrometry then determines their chemical signatures. When a new illicit steroid is found, its chemical signature needs only to be added to those against which the lab results are compared.

But THG eludes the test because the gas-chromatography step does not pull it out. So after determining the molecular structure of the compound in the syringe — which took more than a month — the UCLA lab had to develop a new test for it from scratch. "As far as I know this has never happened before," says laboratory head Don Catlin.

THG may be the tip of the iceberg, says



Running wild? The effectiveness of tests for banned sports drugs has been thrown into question.

Gary Wadler, a sports physician at the New York University School of Medicine and an authority on sports drugs: "It would be hard to believe that this is the only one out there."

There are strong incentives for elite athletes to use illegal performance-enhancers. When everyone is in peak condition, even a marginal benefit can mean the difference between winning and losing. And the more athletes use steroids and other enhancers, the more the incentive grows. "If they think they won't get caught, they do it," says Cynthia Kuhn, a drug-abuse researcher at Duke University in Durham, North Carolina.

Not getting caught means either quitting the supplements weeks or months before a competition at which testing is required, or using something that cannot be detected. Demand for the latter is driving a lucrative

underground market, says Wadler: "There has been a lot of bathtub chemistry going on with steroids for years."

Chemists who develop new steroids have little to fear from the law. Although banned in most sports, the compounds only fall under the US Controlled Substances Act if they are shown to build muscle, Wadler says.

Last year, Catlin's group discovered the synthetic steroid norbolethone in an athlete's urine sample. Developed by the drug firm Wyeth in 1966, it has never been marketed and has no known manufacturer today.

No one knows who first made THG. The USADA has named a Californian laboratory, Balco, as the supplier of the syringe, and has referred the matter to the Department of Justice. Balco denies distributing THG.

Drug-testing labs are left playing a hopeless game of catch-up. Tests occasionally produce signals that correspond to no known compound, says Catlin, but his lab lacks the resources to track them all. "If you decide to go after something, you are committing yourself to a lot of work," he says. ■

Open access wins German support

Quirin Schiermeier, Munich

Germany's main scientific organizations have issued a joint statement backing initiatives that provide free scientific information over the Internet.

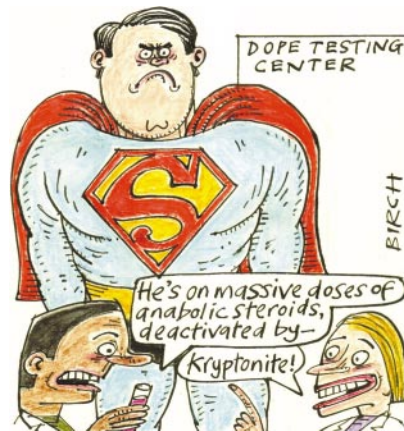
After a three-day meeting in Berlin, organizations including the Max Planck Society (MPS) and Germany's main research-funding agency, the DFG, were due to issue the call for open access on 22 October. Open-access backers say this is the first time that they have won formal support from all major research organizations in a large nation.

The MPS, for example, is changing scientists' employment contracts, requiring them to return the copyright of their work to the society. Researchers will still be able to

publish in scientific journals, but after a grace period — the length of which is still being discussed — their papers must be deposited in at least one online repository.

The declaration evolved from a European Union-funded digital project, European Cultural Heritage Online (ECHO), which facilitates access to cultural materials (see *Nature* 424, 491; 2003).

Robert Schlögl, a chemist at the MPS's Fritz Haber Institute in Berlin, and co-organizer of the meeting, emphasizes that a smooth transition is necessary. "This has nothing to do with confrontation, but it has everything to do with starting a dialogue with publishing houses about a new division of labour," he says. ■



Fish farms' threat to salmon stocks exposed

Quirin Schiermeier

Fears that the Atlantic salmon could become extinct were rekindled last week by a study on the effects of escaped farmed fish on wild fish populations.

The findings are expected to influence legislation in places such as Scotland, where farmed fish are often intentionally released for anglers. In addition, up to two million salmon are thought to escape from farms around the North Atlantic each year (see *Nature* **416**, 571; 2002).

Atlantic salmon are farmed mainly in Norway, Scotland, Ireland, Canada and the northeastern United States, using species bred from a few Norwegian strains for faster growth. This selection process makes the fish more economically valuable, but their unusual genes and their unfamiliarity with natural conditions make them vulnerable to predators and disease in the wild.

The new paper follows a ten-year study carried out in the Burrishoole river system in County Mayo, Ireland, a traditional centre for salmon research run by the Irish Marine Institute (P. McGinnity *et al. Proc. R. Soc. Lond. B* doi:10.1098/rspb.2003.2520; 2003). Researchers released known numbers of farmed salmon, along with eggs of wild-farmed hybrids, in a natural salmon-spawning river. They then recaptured fish in fresh water and at sea, and used DNA profiling to identify their parentage.

As expected, farmed fish were short-lived,



In at the deep end: farmed salmon don't live long in the wild, but their offspring dilute the gene pool.

surviving on average 2% as long as their wild counterparts. But those created by interbreeding became more likely to die in successive generations: hybrids survived 27–89% as long as their wild cousins, depending on their parentage and whether they were first- or second-generation offspring. The second generation fared particularly poorly: 70% of embryos died. “Additive genetic variations may finally cause the extinction of vulnerable populations,” says Andy Ferguson, a fish geneticist at Queen's University Belfast, Northern Ireland, one of the authors of the new study.

The results provide clear evidence of how interbreeding can harm wild salmon, says Eric Verspoor, head of the fish-genetics

group at the Fisheries Research Services' Freshwater Laboratory in Pitlochry, Scotland. It could devastate the beleaguered Atlantic salmon, which has already been hit by pollution and overfishing — wild stocks are extinct, endangered or vulnerable in more than half the 2,000 salmon-spawning rivers around the North Atlantic.

The Edinburgh-based North Atlantic Salmon Conservation Organisation is developing international guidelines for replenishing wild populations. Next year, the group will hold talks with the farming industry about improving containment. Meanwhile, Scottish angling organizations have pledged to alert clubs to the risk of restocking rivers with juvenile farmed salmon. ■

L. CAMPBELL/NHP

Opossum hops over kangaroo to top of genome list

Carina Dennis, Sydney

The humble opossum is set to become the first marsupial to have its genome sequenced — disappointing Australian researchers who want to study the kangaroo.

But the US National Human Genome Research Institute (NHGRI) in Bethesda, Maryland, which will lead the opossum effort, says that if the Australians

raise the money to do half of the kangaroo sequence, it will complete the other half.

The South American grey short-tailed opossum (*Monodelphis domestica*), whose genome is roughly the same size as that of humans, joined the NHGRI's list of high-priority organisms on 14 October. “We expect sequencing to be under way within a year,” says Francis Collins, director of the institute. The tammar wallaby (*Macropus eugenii*), a small kangaroo species, is ranked



Opossum and Paul Samollow: working well together.

as a moderate priority.

Geneticists want a marsupial genome sequence so that they can carry out comparative studies to identify new genes and investigate chromosome evolution. “Mice are too close to humans, and chickens are too far away for comparative genomics — marsupials are just right,” says Jenny Graves, a geneticist at the Australian

National University in Canberra.

“The tammar wallaby is a better choice because so much more is known about it in terms of physiology and genetics,” says Graves. But she adds that she is excited about the possibility of funding for the kangaroo initiative, “as that means we could mount a big genomic enterprise in Australia. It would be a sad state of affairs if we couldn't even sequence our own national emblem.”

The choice came down chiefly to the ease

of working with opossums, researchers say. “There's no comparison — the opossum is much smaller, breeds prolifically and is easy and cheap to keep,” says Paul Samollow, an evolutionary geneticist at the Southwest Foundation for Biomedical Research in San Antonio, Texas, which houses the oldest and largest opossum colony in captivity.

The US kangaroo offer depends upon Australia agreeing to do half of the sequence. An Australian centre for kangaroo genomics was set up in July with a government grant of A\$3.3 million (US\$2.3 million) over five years. But it won't have either the infrastructure or the funding to do the sequencing, which would have to take place at the government-supported Australian Genome Research Facility (AGRF).

“We'll need to raise close to A\$6 million,” says Susan Forrest, director of the AGRF. While hoping for government support, she is also preparing to pass the hat around potential sponsors, including the airline Qantas, whose emblem is a kangaroo. ■

N. GOUN

Biologists seek havens for species at risk

Rex Dalton, San Diego

The survival of threatened species can be substantially improved by creating critical-habitat sanctuaries, where human activities are tightly restricted, says a survey conducted by a US environmental group.

The establishment of such habitats — which is permitted under the 1973 Endangered Species Act — has been a bone of contention between scientists and federal officials for years. Environmentalists have repeatedly tried to use the act to create habitat sanctuaries, whereas the government has argued that such habitats are of little proven value in saving endangered plants or animals.

But a report released on 6 October by the Center for Biological Diversity in Tucson, Arizona, which is involved in several lawsuits to create habitats, says that species recovery is twice as likely to occur in a critical-habitat sanctuary as it is elsewhere. The study is based on an analysis of recently released data, covering 1998–2001 and some 1,300 threatened or endangered species, from the US Fish and Wildlife Service (FWS).

“These findings are very important for the future of endangered and threatened species and are strong support that designation of critical habitat is fundamental for recovery of these species,” says conservation biologist Phil Hedrick of Arizona State



‘Critical habitats’ exist to save threatened species, such as the red-legged frog.

University in Tempe, who reviewed the report at the centre’s request.

A spokesman for the FWS said that the agency had not seen the report, and could not comment on its contents. But he pointed to a FWS position paper from May which states that critical habitats are of “little conservation benefit to species”. Active efforts at conservation are more likely to improve survival than prohibiting activities in habitat sanctuaries, FWS officials say.

The Center for Biological Diversity has charged that the Bush administration was sitting on data released in June for more than two years because they conflicted with the administration’s publicly stated view. But the FWS denies withholding the data, saying the failure to release the material was an oversight.

In Congress, Republicans have repeatedly

and unsuccessfully angled to rewrite sections of the Endangered Species Act, including that on critical habitats. And at least two measures to alter the act are expected to come before Congress next year.

There were only a handful of critical-habitat designations until 1997. But now there are about 435 US critical habitats, covering 38 million acres and 542 species. A further 350 habitat projects are in the works, with about 60 scheduled for completion in the next year.

If a sanctuary is designated as a critical habitat, any federal agency planning actions related to that area must consult with the FWS on potential impacts. The consultations often lead to controls on activities such as logging, mining and road-building.

Kieran Suckling, the centre’s executive director, says that the Bush administration “is starving the FWS budget to slow the protection of critical habitats” ordered by judges. The federal government budgeted \$9 million this year for creating critical habitats. In the past, the FWS has recommended spending up to \$20 million a year for six years to create habitats already designated as critical.

The centre’s report shows that protecting critical habitats is the best way “to move from species-to-species conservation to real ecosystem protection”, says Suckling.

♦ www.sw-center.org

Standards-lab staff up in arms over military link

Jim Giles, London

The contract to run Britain’s National Physical Laboratory (NPL) is up for grabs — and scientists there are warily eyeing the prospect that the nation’s main standards and measurement laboratory could fall under the auspices of a contractor that specializes in military technology.

QinetiQ, which was spun off from the Defence Evaluation and Research Agency in 2001 to run some of Britain’s weapons-development labs, is one of three contenders to run the NPL when its operating contract is renewed in six months’ time.

The NPL, which is based in Teddington near London, had its operations contracted out in 1995 to Serco, a services company that runs everything from sports centres to government computer contracts. But Serco’s contract runs out in April 2004, and the Department of Trade and Industry (DTI) has invited bids from companies to run the

lab until 2014. QinetiQ, Serco and Scientific Generics, a Cambridge-based consultancy, are all bidding for the contract.

Serco already has some defence interests, but trade unions at the NPL say they find QinetiQ’s military focus incompatible with the lab’s civilian tradition. “Lots of scientists at the NPL made a conscious choice not to go into defence research,” says Fiona Sloman, who represents more than half of the 600 staff at the laboratory for the union Prospect. “They would be concerned about QinetiQ taking over,” she adds.

Government officials say that whoever wins the contract will have to pursue the same research agenda, drawn up by the laboratory and the DTI.

But laboratory staff say they are worried that QinetiQ’s involvement might deter researchers from joining the lab. “I have big problems with defence companies,” says one physicist who joined NPL in part because it



War machine: QinetiQ is known for its defence research, such as the design of this plastic tank.

was an applied lab that did not have links to weapons research.

In early November, the DTI is expected to announce which of the three applicants it will choose as its preferred bidder, with whom it will then enter detailed contract negotiations.

♦ www.npl.co.uk

Court ruling sees ban on medical marijuana go up in smoke

Washington Doctors in the United States can continue to recommend marijuana as pain relief for their patients, thanks to a Supreme Court decision last week.

Most of the United States takes a hard line against the use of marijuana, but some states — including California, Alaska, Arizona, Colorado, Hawaii, Oregon, Maine, Maryland, Nevada and Washington — allow the drug to be used for medicinal purposes.

Since 1996, when California passed the first law to legalize medicinal marijuana use, the US government has threatened to take away doctors' rights to prescribe any

federally controlled substance if they recommend marijuana to their patients. In 2000, the 9th Circuit Court of Appeals in San Francisco barred the US government from acting on its threat. The Bush administration challenged the injunction last year, but last Tuesday the Supreme Court refused to hear the case.

The decision has been hailed as a major victory by supporters of medical marijuana, who say that the drug is often the only way for patients with diseases such as cancer and AIDS to control pain and nausea.

Computer programs to call the shots for astronomers

Paris Astronomers could soon be alerted to events in the skies by text messages to their mobile phones, after a successful trial of new computer technology.

The service will come courtesy of computer programs called agents, developed as part of Britain's eScience Telescopes for Astronomical Research project. The agents are designed to collect data from sky surveys, sift through them for anything unusual, and then decide what to do. If an agent detects a pattern that matches a short-lived event such as a γ -ray burst, for example, it can rapidly alert astronomers to the event and direct telescopes towards it.

The agents take advantage of the Grid — a system that stitches together computing power from around the world (see *Nature* 422, 799–800; 2003) — to help crunch vast amounts of data, and to communicate with each other and the telescopes.

In tests on the UK Infrared Telescope on Mauna Kea in Hawaii, reported last week at a meeting in Strasbourg, France, one agent detected a dwarf nova — a star that shows flares in brightness. The network will now be expanded to include other telescopes.

Radiation lab brings space studies on-beam

Washington Six years after it was proposed, NASA finally has a dedicated facility to study the biological effects of radiation — considered by many to be the biggest barrier to sending humans into deep space (see *Nature* 389, 320; 1997).

The \$34-million NASA Space Radiation Laboratory was opened last week at Brookhaven National Laboratory in New York state. Beams of heavy ions from Brookhaven's Booster accelerator will be used to bombard tissue samples, rats and other biological samples with radiation similar to what they would experience outside Earth's magnetic field. The beams, which are shared with physics experiments,



Joint effort: US doctors have won the right to talk to patients in pain about using marijuana.

can also be used to test radiation shields.

NASA's chief of staff, John Schumacher, says that although some radiation research will still need to be done in space, 80% of it can now be conducted on the ground. Such studies are critical to space exploration, he says. "If we don't succeed, we're not going anywhere. That's how critical this work is."

Laser satellite blows wind of change for weather studies

London Measuring wind speeds from space could improve forecasts of hurricanes and tropical storms, say researchers involved in developing an experimental satellite.

Winds are usually monitored by land-based stations, balloons and aircraft. But Martin Jones, of the UK Meteorological Office in Bracknell, near London, says the existing network provides an incomplete picture. Jones and others hope that the satellite — named *Aeolus* after the character who controls the wind in Homer's *Odyssey* — will help fill the gap. *Aeolus* will scan the sky with a laser and determine wind speeds by monitoring changes in the reflected beam.

The contract to build *Aeolus* is due to be signed this week by the European Space Agency and EADS Astrium, a company based in Stevenage, UK. The satellite is scheduled for launch in 2007.



Eye on the storm: weather experts hope to use satellite data to forecast hurricanes.

Cash boost gets Earth project off the ground

Washington The Earth sciences received a major boost last week as the US National Science Foundation (NSF) agreed to spend \$219 million over the next five years on EarthScope, a suite of observation projects.

The programme will use satellites and a network of 400 seismometers to improve geophysical monitoring of the planet (see *Nature* **405**, 390–392; 2000). The sensors will map and monitor, in real time, a wide range of natural phenomena, including

plate tectonics, earthquakes and volcanic eruptions. They will also gather data from beneath Earth's surface, to study the planet's magnetic field and the motion of the continental plates. Some EarthScope projects are already under way, as the NSF has allocated some \$30 million to the programme in 2003.

Water batteries make a splash in electric field

London Engineers in Canada have found a means to generate electricity by wringing it out of water.

Conventional hydroelectric generators harness energy from water flowing downhill. But Larry Kostiuk and his colleagues at the University of Alberta have figured out how to use the electrical charges in water itself. In their 'electrokinetic battery', water is forced through micrometre-sized pores in a glass filter.

Because the glass surface is negatively charged, the positive ions in the water are drawn through in preference to the negative ones. The device can generate a current of 1.5 microamps.

The catch is that the electrokinetic battery is so far only about 0.1% efficient — large-scale hydroelectric generators boast efficiencies of about 80%.

The rough guide to the genome



A new effort to map human genetic variation should provide a shortcut for researchers trying to uncover the roots of disease. Carina Dennis profiles the 'HapMap' project.

Searching for the genetic factors associated with common ailments such as heart disease and asthma is like trying to find a needle in a haystack. But an international consortium is now working on a way to speed the search, by packaging the haystack of human genetic variation into more manageable bundles.

The three-year, US\$100-million endeavour, called the International HapMap Project, is the biggest thing in genomics since the effort to sequence the entire human genome, which was completed in April this year. If that sequence is the 'Book of Life', then the HapMap will be a handy index, helping gene-hunters to zoom in quickly on informative chapters or pages. "The HapMap will provide the missing link between the DNA sequence of the genome and the way in which the genome influences the risk of disease," says Francis Collins, director of the National Human Genome Research Institute in Bethesda, Maryland, one of the HapMap's main financial backers.

Geneticists want to know why some of us are more likely than others to succumb to a particular disease. But they often have no idea where to start looking among the three billion letters, or bases, of the human genetic code. So researchers often sample populations that are affected by the disease, and study single nucleotide polymorphisms (SNPs) — points in the genome at which one base can differ between individuals. If a particular SNP is inherited with the disease, it is a strong indication that a gene that confers susceptibility lies somewhere nearby.

There are thought to be some ten million common SNPs scattered throughout the human genome. Testing thousands of people for millions of SNPs is beyond current technology — so in practice, gene-hunters use a few thousand SNPs at most. And with no prior knowledge of which SNPs will be most useful, it is a hit-and-miss business.

The HapMap project aims to use our understanding of the genetic shuffling that occurs during the production of sex cells to make gene searches more efficient. In the cell divisions that give rise to eggs or sperm, pairs of chromosomes line up and exchange portions of genetic material. This is not entirely random: the breaking, exchange and reas-



On the road: Francis Collins launches the HapMap project.

ing tend to occur at particular points. As a result, blocks of sequence have been inherited down the generations from our common ancestors without being broken up. The blocks vary widely in size, but most are thought to be 10,000 bases or more in length¹.

The DNA sequence of a block is known as a haplotype. Common haplotypes can be identified by sampling only a handful of key SNPs, and studies have indicated that, for each block, there are only a few common haplotypes in the population¹⁻³. The HapMap should therefore provide geneticists with the means to scan the entire genome rapidly for disease genes, perhaps by analysing as few as 300,000 SNPs¹. The map could also be used to study how people respond differently to prescription drugs, according to their particular genetic make-up (see page 760).

Like the publicly funded effort to sequence the human genome, the HapMap consortium has a commercial rival. The HapMap team has also made sure to consult with the communities from which DNA will be sampled, to avoid the charges of scientific 'colonialism' that have dogged earlier attempts to study human genetic diversity.

Three-pronged attack

The consortium will focus on DNA samples collected from three distinct populations: people of northern- and western-European descent; a Nigerian population called the Yoruba; and an Asian collection composed of Japanese and Han Chinese. Analysis of the samples from people of European ancestry, which were collected for earlier population-genetic studies, is already under way.

The HapMap consortium will produce a haplotype map based on the similarities in block structure and common haplotypes for the European, African and Asian populations. "The block architecture will be similar, but the common haplotypes are likely to vary between different populations in terms of composition and frequency," says Pui-Yan Kwok, a geneticist at the University of California, San Francisco, and a member of the HapMap consortium.

The idea is not to document the full extent of human genetic diversity, but to provide useful tools for gene discovery. If



it becomes obvious that other populations possess haplotypes that are not represented in the maps, this information can be incorporated at a later date. The Yoruba population, however, is expected to contain the greatest genetic diversity. "What you find in European and Asian populations, you almost always find in Africa," says Collins.

With its strong biomedical focus, the HapMap project is quite different from the Human Genome Diversity Project, an ill-fated exercise first proposed more than a decade ago. That project's founders were interested in anthropological questions. They wanted to sample DNA from many populations to understand their genetic relationships to one another, and to link the findings to studies of language and other cultural phenomena. Many groups representing indigenous peoples objected, arguing that it would exploit vulnerable populations and intrude into people's beliefs about their own origins. Fatally wounded by these accusations, the project never got off the ground.

Given the potential for controversy, the HapMap consortium's ethical experts are going to great lengths to obtain informed

TOP: D. HALLMAN/STONE; ABOVE: P. ALMAMY/CORBIS



World view: the HapMap initiative will gather genetic data from African, Asian and ancestrally European populations.

across the genome at intervals of around 5,000 bases. Further SNPs will then be identified where needed to define haplotype blocks. The team is also running a pilot study in parallel to scrutinize every known SNP in ten selected regions of the genome, each some 500,000 bases long. This should reveal whether the map can be produced using fewer SNPs, or whether more will be required.

In keeping with the precedent set by the Human Genome Project, the HapMappers will release their data as soon as they are available. "That way, other people can work on the data at the same time and come up with new ideas for analysing them," says Peter Donnelly, a statistician at the University of Oxford, UK, and a HapMap collaborator.

Patent parasites?

But this policy of openness has led to concerns about what Collins calls "parasitic intellectual property claims". The project's leaders fear that some people could combine the consortium's data with their own, and then patent the findings in a way that restricts others' ability to work freely with the HapMap data.

Earlier rushes to patent SNPs fell by the wayside when it became clear that patent offices did not consider them to pass the test of having a clear 'utility'. But as knowledge of genetic variation grows, some may claim that haplotypes are patentable tools for gene discovery. "Haplotypes, more than SNPs, come closer to patentability," says Lawrence Sung of the University of Maryland School of Law

in Baltimore, who is advising the HapMap consortium. So users of the HapMap database, run by the Cold Spring Harbor Laboratory in New York state, will be required to click on an agreement to certify that they will not take any action that would exclude others from using the data.

Whether or not haplotypes are patentable, one company is trying to turn them to profit. Perlegen, based in Mountain View, California, claims already to have made its own haplotype map. Two years ago, the company unveiled its data for human chromosome 21 (ref. 3). Pharmaceutical firms including GlaxoSmithKline and Bristol-Myers Squibb have now enlisted Perlegen to sift through their clinical DNA samples to track down genetic markers that might explain why individual patients respond differently to drugs.

The relationship between Perlegen and the HapMap consortium has so far been more cordial than competitive. And both sides are adamant that the company's work does not render the public effort redundant. "Often the only way we know a scientific result is right is when multiple people can arrive at the same answer," says David Cox, Perlegen's chief scientist.

Flustered gene-hunters whose quarries remain hidden in the vastness of the human genome can only hope that the HapMap will help to answer their prayers. ■

Carina Dennis is Nature's Australasian correspondent.

1. Gabriel, S. B. *et al. Science* **296**, 2225–2229 (2002).
2. Dawson, E. *et al. Nature* **418**, 544–548 (2002).
3. Patil, N. *et al. Science* **294**, 1719–1723 (2001).
4. Reich, D. E., Gabriel, S. B. & Altshuler, D. *Nature Genet.* **33**, 457–458 (2003).
5. Carlson, S. C. *et al. Nature Genet.* **33**, 518–521 (2003).

► www.hapmap.org



consent from the study populations, and to involve communities in decision-making. They have given extensive thought to the question of how to explain the project's rationale in language that ordinary people can understand. All DNA donors will be asked to give consent for their samples to be used not just for the HapMap project, but also for future studies of genetic variation. For each sample population, a community advisory group will be set up to ensure that these future studies are consistent with the consent form.

Other challenges are technical — the first being to work out which of the millions of SNPs identified so far will be most useful in constructing the maps. Some SNP variants, for instance, occur at frequencies that are too low to be readily studied in the sample populations^{4,5}. The HapMappers also need to ensure that they have a good, even coverage of SNPs across the entire genome. So initial work on the project, which began in October last year, has concentrated on addressing these issues. "We have beefed up SNP density almost threefold across the genome," says Collins.

The consortium plans to construct the initial map by studying 600,000 SNPs scattered

With your genes? Take one of these, three times a day



Truly 'personalized' medicine remains a distant goal. But researchers are now thinking about how to use genomic data to avoid prescribing drugs that may kill, or won't work. Alison Abbott reports.

It causes at least 100,000 deaths each year in the United States, and is responsible for more than 10% of hospital admissions in some European countries. But this isn't some terrifying emerging infection, it's the annual toll inflicted by adverse reactions to prescription drugs. What's more, millions of people are being treated with drugs that, for them, will never do much good. Beta-blockers, given to reduce blood pressure, are ineffective in one-third of patients; many antidepressants don't work in half of the people who take them.

The blame lies largely with our genes, which help to determine the way in which we react to drugs. Small genetic variations

between people — or polymorphisms — can alter the behaviour of proteins that carry a drug to its target cells or tissues, cripple the enzymes that activate a drug or aid its removal from the body, or alter the structure of the receptor to which a drug is supposed to bind. Variation in immune-system genes can also influence how particular drugs are tolerated. Together, these subtle genetic variations mean that the dose at which a drug will work may vary hugely from person to person. And with today's 'one-size-fits-all' prescribing, that can lead to life-threatening adverse reactions or to a drug completely failing to do its job.

Yet the genomics revolution has given us the tools to identify people who don't fit the

standard prescribing mould. Single nucleotide polymorphisms, or SNPs, are single-letter changes in the genetic code that are scattered throughout the genome. They can now be profiled with increasing efficiency, and used to highlight polymorphic genes that influence our response to individual drugs. Will we see a seismic shift in prescribing practice? With barely a handful of concrete examples to go on so far, it's too early to judge the true promise of personalized medicine — but the pharmaceutical ground is certainly starting to shake.

Regulatory agencies such as the US Food and Drug Administration (FDA) are starting to consider whether or not some drugs should be labelled as being suitable only for individuals with a defined genetic profile. Next month in Washington DC, the FDA will hold a workshop on the topic. Drug companies are preparing themselves for a new lie of the land. Already, three-quarters of the clinical trials they submit to the FDA for approval include provision for sampling and storing blood for any genetic analysis that may be required in the future. But it is academic researchers who are most enthusiastically pushing things forward.

Pharmacogenetics — the study of the influence of genetic variation on drug responses — isn't a new discipline. But until the advent of SNP analysis, progress was laborious. An early example of the field's promise came with the introduction, in the 1950s, of the muscle relaxant succinyl-



Off target? The 'one-size-fits-all' approach to drug prescription ignores genetic variations that can cause side effects or poor response in some people. Simple blood tests may offer a more bespoke service.



choline. To their horror, anaesthetists found that a small proportion of patients given the drug went into life-threatening respiratory arrest on the operating table. Geneticists were able to show why. Succinylcholine is normally metabolized very efficiently by an enzyme called cholinesterase. But 1 in 2,500 people carry two defective copies of the gene for this enzyme, and suffer deep and prolonged paralysis of muscles, including those needed for breathing, when given the drug.

Hidden from view

Being rare, such reactions are unlikely to be noticed in clinical trials, which typically involve a couple of thousand patients. They tend to be revealed when a drug is widely prescribed — which is why regulatory agencies require drug companies to supply information about adverse drug reactions for years after approval. If such reactions occur in an identifiable group of patients, the drug's labelling can be modified to exclude them. In the case of succinylcholine, bad responders are identified by measuring cholinesterase activity in the blood.

The FDA is now close to making a landmark decision to relabel an approved drug to exclude bad responders identified by SNP analysis alone. Mercaptopurine, approved by the agency in the 1950s, is used to treat childhood leukaemias and other cancers. Like all anticancer agents, it is intrinsically toxic, and



Analytical express: genotyping of samples has become a highly automated business.

there is a narrow dose range within which the therapeutic benefit outweighs its toxicity. But some patients suffer life-threatening bone-marrow damage at doses tolerated well by others. William Evans, scientific director of St Jude Children's Research Hospital in Memphis, Tennessee, suspected that this could be linked to polymorphism in the gene for the enzyme thiopurine methyltransferase, which metabolizes the drug.

In the mid-1990s, Evans began to compare the sequences of this gene in patients with and without the toxic reaction. He identified three SNPs, any of which resulted in an enzyme that allowed mercaptopurine to linger in the body at dangerous levels¹.

"Altogether we found that 70% of cases of toxicity have carried one or more mutations," says Evans. "SNP testing gives clinicians a much more robust and easy-to-handle method of estimating this risk than measuring activity of the enzyme." Even before the FDA has given its seal of approval, some hospitals are now subjecting patients to genetic tests before prescribing mercaptopurine.

Most research in this field has similarly focused on polymorphisms affecting drug metabolism. Pharmacogeneticist Magnus Ingelman-Sundberg of the Karolinska Institute in Stockholm, Sweden, estimates that the efficacy or safety of 20% of the drugs now on the market is associated with a polymorphism in a single gene for a metabolizing enzyme. A similar proportion may be influenced by polymorphisms in multiple drug-metabolizing genes. Factor in the largely unknown quantity of polymorphisms influencing the transport of drugs or their interaction with receptors, plus immune-system genes that influence tolerance to drugs, and it's clear that pharmacogeneticists have barely scratched the surface.

Family fortunes

One key group of drug-metabolizing enzymes is the P450 family. These enzymes are produced in the liver and oxidize foreign chemicals. Of the 57 genes for P450 enzymes that have been identified in humans, three — *CYP2D6*, *CYP2C9* and *CYP3A4* — are particularly important for drug metabolism. *CYP3A4* metabolizes half of all prescribed drugs but seems to have few polymorphisms that could affect drug function; the polymorphic *CYP2D6* is involved in the metabolism of a quarter of prescribed drugs — including beta-blockers and antidepressants; *CYP2C9* is involved in the metabolism of 5% of drugs and is also highly polymorphic².

Although many of these polymorphisms will have no effect on enzyme function, some have already been linked to the failure, in certain patients, of commonly prescribed drugs. For example, the painkiller codeine depends for its effect on being oxidized to morphine by *CYP2D6*; people with a particular polymorphism — up to 10% of the population — don't get pain relief. The hit-and-miss efficacy of antidepressants such as Prozac may also be partly the result of *CYP2D6* polymorphisms³.

Other P450 polymorphisms are associated with serious adverse drug reactions. For example, the widely used anticoagulant warfarin is metabolized by the enzyme encoded by *CYP2C9*, and the small proportion of patients with polymorphisms that reduce the enzyme's activity can suffer life-threatening bleeding unless warfarin doses are reduced⁴. In one dramatic case in the 1980s, a drug for angina called perhexiline had to be pulled off the market when it turned out to be toxic to liver and nerve cells in rare cases of a particular *CYP2D6* polymorphism. Ingelman-Sundberg estimates that some 80% of the most



serious adverse drug reactions involve drugs metabolized by polymorphic P450 enzymes. "We can learn a lot about how to prescribe to patients by thinking about an individual's drug-metabolizing enzyme profile," he says.

With this in mind, several genomics companies are manufacturing DNA microarrays to identify common SNPs that influence the activity of P450 enzymes, and some hospitals are starting to use them. In the long run, such chips could help not just to avoid dangerous reactions to drugs, but also in deciding appropriate drug doses, and selecting which drug cocktails to give to patients with complex conditions such as cardiovascular disease.

Profiles on parade

Given the promise of pharmacogenetics, funding agencies are starting to invest. In June, the UK Department of Health earmarked £4 million (US\$6.7 million) over three years for pharmacogenetic research into existing drugs, and in 2001 the US National Institutes of Health (NIH) established a Pharmacogenetics Research Network through which 12 partners across the country, including Evans, will share their experiences and post results of large studies on a public website, to encourage further research. These studies include the investigation of polymorphisms in neurotransmitter receptors relevant to antidepressant therapy, and those involved in cell-signalling pathways that help to regulate blood pressure.

Another member of the NIH pharmacogenetics network is Jeffrey Drazen of Harvard Medical School in Boston, editor-in-chief of *The New England Journal of Medicine*. An increasing number of papers submitted to the journal are relating treatment outcomes to particular genetic profiles, but Drazen notes that academic studies tend to be small, which limits their statistical power.

Larger studies will be needed to determine the influence of groups of SNPs on drug responses in diseases, such as cardiovascular and psychiatric conditions, which are influenced by multiple genes. This will probably involve searching for informative SNPs across the entire genome — an expensive process that would mark the transition from conventional pharmacogenetics to the era of high-throughput pharmacogenomics.

Such projects may require the financial muscle of the drug industry. But for now, most companies are reluctant to get involved: their priority is bringing lucrative new drugs to the widest possible market, rather than limiting their use to patients with particular genetic profiles. "Our general philosophy is not to initiate a drug-development programme that would limit the group of patients a drug could treat," says Brian Spear, director of pharmacogenetics at Abbott Laboratories in Illinois.

Drugs firms also worry that genomic information could be interpreted with excessive caution by regulatory officials — keeping



Allen Roses has identified genetic polymorphisms that cause severe side effects in an AIDS therapy.

useful drugs off the market. Larry Lesko, director of the FDA's office of clinical pharmacology and biopharmaceutics, is trying to address these worries. "Two or three years ago, we became aware of concern from companies that wanted to integrate pharmacogenetics into their drug-development programmes but were afraid we would overreact," he says.

Lesko created the concept of a 'safe haven' for genomic data. This allows information to be deposited so that, for instance, companies can make claims about a drug's efficacy in a defined population. Until better methods of analysing the data are developed, the FDA won't use this information to make judgements about safety. The concept is still being refined, but a proposal will be discussed at the November FDA workshop. "We hope that it will make companies feel freer to explore pharmacogenomics," says Lesko.

Blood banked

Most companies have seen the writing on the wall, and now store blood samples from clinical trials in case they need to analyse them retrospectively. But they want to see methods for pharmacogenomic analysis become more reliable before routinely making them part of their clinical trials. "We still don't know how predictive SNPs will be in complex diseases where lots of different factors may compensate for the effect of a polymorphism," says Klaus Lindpaintner, director of Roche Genetics in Basel, Switzerland. "Another, sometimes overlooked, problem is lack of good clinical data that show how SNPs relate to clinical responses."

Against this conservative backdrop, Allen Roses, senior vice-president of genetics research with GlaxoSmithKline (GSK) in Research Triangle Park, North Carolina, stands out as an enthusiast for SNP analysis. "Ten years down the road, the public will be insisting that even suppliers of vitamin supplements are required to supply information on polymorphisms," he predicts. GSK has several pharmacogenetic projects running, some of which involve scanning the whole genome for thousands or tens of thousands of

SNPs using various technologies. "This is simpler and cheaper than most people imagine," says Roses.

It will be years before all of these studies will be completed and published, but Roses has won praise from researchers for an initial foray into pharmacogenomics, published last year in *The Lancet*⁵. His team went through the GSK database of clinical trials to identify 85 patients who had taken the AIDS drug abacavir and experienced a violent immune reaction to it. By comparing hundreds of SNPs from these patients with those from others who did not suffer the life-threatening side effect, Roses identified three SNPs associated with immune-system genes that can be used to identify patients who shouldn't be given abacavir.

Trial-busters

Most companies believe that pharmacogenomics could offer a way to increase the efficiency of clinical trials. In early-phase trials, a new drug is tested on about 250 patients to see if it causes toxicity and to see whether it works. On the basis of these results, regulatory authorities then decide whether to give the go-ahead for larger 'phase III' studies. If a particular genotype can be distinguished that is associated with a good response to a drug, then patients with such a genotype could be selected for the phase III trial. This could allow companies to recoup their investment in candidate drugs that ultimately prove to be effective only in a proportion of people. The FDA is considering this, but Lesko points out that safety studies would still have to be carried out in non-responders, which could mean expanding earlier phases of the trials.

As the potential of pharmacogenomics becomes clearer, its application by both drug companies and health services will be dictated by balancing costs and benefits. How many SNPs will be needed to make accurate predictions for each drug and condition, and how much will it cost to analyse them? Will the cost be worth it if the clinical benefit is small?

Those won't be easy questions to answer. But Drazen believes that the benefits of pharmacogenomics will be visible in the clinic within a decade. Currently, he notes, a doctor faced with a depressed patient has to choose from a long list of drugs, all of which work only in some patients, and carry a high risk of undesirable effects. "In future, that physician would send off a blood sample for analysis and get back a ranked list of best options for that individual patient," says Drazen. For patients who are subjected to the current prescribing lottery, that day can't come soon enough. ■

Alison Abbott is Nature's senior European correspondent.

1. Krynetski, E. Y. & Evans, W. E. *Oncogene* **22**, 7403–7413 (2003).
2. Pirmohamed, M. & Park, B. K. *Toxicology* **192**, 23–32 (2003).
3. Arranz, M. J., Mancama, D. T. & Kerwin, R. W. *Curr. Pharmacogenomics* **1**, 151–158 (2003).
4. Higashi, M. K. et al. *J. Am. Med. Assoc.* **287**, 1690–1698 (2002).
5. Hetherington, S. et al. *Lancet* **359**, 1121–1122 (2002).

Bioethics needs a distinct voice if it is to aid science

Neither scientific knowledge nor gut feeling is enough to support decision-making.

Sir — Paul Copland in his Correspondence “Science and ethics must not be separated” (*Nature* **425**, 121; 2003) argues that ethics is an integral part of science and therefore the two must not be separated. But I believe that we should distinguish between separation (which is necessary to allow for the development of expertise) and the use of unscientific approaches (which can make communication with scientists impossible).

Scientists should therefore not argue against the establishment of the philosophical discipline of bioethics; rather, they should welcome it. What science should argue for is a ‘scientific’ rather than a dogmatic approach in the humanities, and specifically in bioethics.

Philosophically, perhaps the biggest achievement of science is the abandonment of dogmatism and the acceptance that all scientific knowledge can potentially be changed by new data and new insights. Similarly, scientists should request from scholars in the humanities that they abandon any “ill-defined ‘personal philosophy’ and ‘gut feeling’” and open themselves to an informed search for the better argument.

It is naive to believe that the views of scientists (especially those directly involved in discoveries) would be more objective

than those of expert bioethicists.

The problem with science is that we often do not know the answers for certain. Will stem-cell research develop a cure for Parkinson's disease? Will transgenic crops pose any risk at all? And therefore we can only present data, which then form the basis of non-scientific decisions. There is no reason to assume that scientists would be better equipped to make these decisions than non-scientists.

The fact that bioethics has developed as a discipline distinct from science simply

reflects the reality that, unfortunately, most scientists do not have the time to become experts in the philosophy of science and bioethics.

We should therefore not fight the development of a philosophical discipline of bioethics, but should ensure that the approaches the discipline takes are scientific and undogmatic.

Alfons Lawen

Department of Biochemistry and Molecular Biology, Building 13D, School of Biomedical Sciences, Monash University, Victoria 3800, Australia

Bioethics: role of religion cannot be ignored

Sir — I agree with P. Copland, who wrote in Correspondence (“Science and ethics must not be separated” *Nature* **425**, 121; 2003) that as scientists we are in a “privileged position” to acquire and interpret scientific information and its ethical implications. But with this privilege comes the responsibility to listen carefully to the concerns of the wider community, whether its members understand developmental biology or not. Otherwise, we risk marginalization in the bioethics debate.

As scientists we are clever enough to know that our articulation of scientific

ingenuity to non-scientists increases cash flow for biomedical research. To insist, as Copland does, that these same benefactors are unqualified to grapple with complex bioethical issues is incorrect.

Societal ethics cannot be conveniently separated from religion when most members of society derive their ethics from religious sources. Indeed, whether scientists think such sources are important or not is completely irrelevant; the religious contribution to ethical debate cannot simply be ignored.

Stephen J. McSorley

University of Connecticut Health Center, Department of Medicine, Division of Immunology, 263 Farmington Avenue, Farmington, Connecticut 06030-1319, USA

Unnatural coverage of Chernobyl tragedy

Sir — Your Books and Arts exhibition review “Unnatural causes” (*Nature* **425**, 347; 2003) described two photographic exhibits at the Tenth International Biennale of Photography in Turin, Italy. These included photographs by the Kiev artist Ilya Chichkan of deformed fetuses that he “borrowed” from mothers living in Kiev during and after Chernobyl. Chichkan removed the fetuses from their jars of formalin, dressed them in jewels and photographed them. The review was accompanied by one of the photos it described.

I understand that Chichkan may have been trying to emphasize the tragedy of Chernobyl. However, I found both the exhibit and your description of it to be gratuitous and tasteless. Rather than seeing a “normal sleeping child”, I saw a tragically deformed fetus, killed by the carelessness of men, and perversely dressed up like a doll. Rather than seeing one of the “sleeping princes of Ukrainian legend” in

“anachronistic dignity”, I saw a poor dead child exploited for media sensationalism. Your coverage of this exhibit is what I would expect from a lurid tabloid, not a pre-eminent science journal. In the future, please remember that although many of *Nature's* readers are biologists, we are also mothers and fathers, with parents' sensibilities.

Oliver R. W. Pergams

Department of Conservation Biology, Chicago Zoological Society, Brookfield, Illinois 60513-1095, USA

India's mission to the Moon is worthwhile

Sir — Your New Delhi correspondent attributes views to me (*Nature* **424**, 985; 2003) that imply I do not support the Moon mission of the Indian Space Research Organisation (ISRO).

As I had no discussion with your correspondent just before the report was filed, I suspect that it was based

on comments I made about a year ago to the effect that “exploration of the Moon should really be an international enterprise, and India, as one of the countries that has the capability to undertake such an exploration, should naturally be a part of that enterprise”. This is a view I still hold, and is actually an endorsement of India's joining the enterprise, especially as 15% of the payload on the Indian satellite to the Moon is available for international collaboration.

Similarly, my belief that exploration of our own planet is of profound importance, should not be interpreted as opposition to the Moon mission. Indeed, I believe that the Moon project conceived by the ISRO is both feasible and worthwhile.

Roddam Narasimha

National Institute of Advanced Studies, Bangalore 560 012, India

correspondence

Contributions to Correspondence may be submitted to corres@nature.com. They should be no longer than 500 words, and ideally shorter. Published contributions are edited.

Outstanding in his field

James Clerk Maxwell's work on the electromagnetic field started a revolution.

The Man Who Changed Everything: The Life of James Clerk Maxwell

by Basil Mahon

Wiley: 2003. 254pp. £18.99, \$27.95

John Maddox

James Clerk Maxwell was the most perceptive physicist in the interval between Faraday and Einstein, H. A. Lorentz, Heinrich Hertz and Ludwig Boltzmann and their like notwithstanding. Basil Mahon, the author of this affectionate biography of Maxwell, is an ex-officer in the British Army and a former civil servant, who confesses that at the age of 16 he had "fallen under the spell" of Maxwell. Half a century after the infatuation, the book is evidently a retirement project. But it is none the worse for that.

Mahon tells the tale of Maxwell's childhood economically but well. Maxwell's father had been an unenthusiastic Edinburgh lawyer who married in his late thirties and, with his new wife, embarked on revitalizing a 6-square-kilometre estate that he had inherited in Galloway in southwest Scotland. James was born in 1831 and was educated at home by his mother, Frances, until her death when he was eight (after surgery for cancer without the benefits of anaesthetics). Two aunts in Edinburgh then persuaded the widower that young James should attend a proper school — the Edinburgh Academy.

Maxwell's first scientific paper — an ingenious generalization of the familiar construction of an ellipse by anchoring an inelastic string to two fixed points and letting a pencil constrained by the string trace out a curve — was published while he was still at school, aged 14. That ensured he was noticed in the small-town Edinburgh of 1846.

Father and son were still undecided about James's future, so he signed up for the standard degree course at the University of Edinburgh and immersed himself in science and mathematics. There he published two further papers, one of which — a mathematical framework for the treatment of the distribution of stress and strain in an elastic solid — became, mostly in other hands, one of the monuments of British nineteenth-century physics.

After three years, his father abandoned his ambition that his son should become a lawyer. James's scientific bent was acknowledged and he was sent off to Cambridge, where he prospered marvellously. He graduated in 1854, at the age of 23, as 'second wrangler' (he was placed second in the lists of those completing the mathematical Tripos examination) and joint holder of that year's



ILLUSTRATIONS BY AL GRANT

Smith's Prize. An academic career, beginning with a fellowship at Trinity College, Cambridge, was assured. And Maxwell began in fine style, by establishing the three-colour theory of colour vision: the assertion that red, green and blue light in an appropriate mixture will give any colour.

His formal academic appointments were all brief. He was professor of physics at the Marischal College until made redundant by the merger that created the University of Aberdeen; he held the same post at King's College London for five years; and, after a gap of six years, he was professor of experimental physics at Cambridge, where he oversaw the design, construction and commissioning of the Cavendish Laboratory until his untimely death seven years later.

Mahon repeatedly reminds his readers that Maxwell was not a mere theoretician but could devise (and then carry out) telling experiments. There were, for example, spinning tops carrying differently coloured paper on whose colour subjects

would be asked to comment. But Maxwell's high point as an experimentalist must surely have been his design of equipment to measure the viscosity of a gas by means of counter-rotating discs in a closed cylinder — a crucial test of his kinetic theory of gases — which was installed and operated in the attic of his London house.

His main body of work was virtually complete within 18 years of his graduation. It consisted of his kinetic theory of gases and his theory of the electromagnetic field, better known as Maxwell's equations. Mahon has also a salutary reminder that Maxwell, as chairman of a committee appointed by the British Association to make sense of electrical units, provided the whole of science with a rational way of choosing units of measurement. How could Maxwell have been so prolific, and in such a short time?

The simple answer is that Maxwell was a genius, but that is a cop-out. What kind of genius was he? First, his penetrating insight into the nature of reality was primarily geometrical. His elegant use in an 1867 paper of

a simple diagram to relate the velocities of two molecules before and after a collision makes that plain.

Second, he was not a perfectionist. His algebra, for example, was often spattered with simple errors, unworthy of even a second wrangler. And he was not afraid to publish theories that were manifestly incomplete. Thus Maxwell's equations (published in 1872) were preceded by an attempt to put Faraday's lines of force into mathematical language (in 1861) and by a fanciful model (in 1865) in which the whole of space was occupied by tiny cells capable of carrying charge and rotation (whence the magnetic part of the field). Mahon performs the public service of providing an understandable account of the workings of the tiny cells that were supposed to be the basis of the electromagnetic field.

Third, Maxwell kept worrying away at problems (the kinetic theory and electrodynamics) until their solution had the ring of truth — and then he wrote a book on the subject.

Maxwell's equations were famously found to contain the conclusion that electromagnetic waves travel with the speed of light. Maxwell (and his contemporaries) saw what a breathtaking result that was. Sceptical of action at a distance, the cell-like structure with which he had endowed all space was replaced by the doctrine of the lumeniferous ether.

It seems that Maxwell fell in with the fashion of the day, even to the extent of suggesting in a letter to the director of the US Naval Observatory that the expected 'ether drift' (in which the ether is partly carried along by massive objects such as planets) might be measured by accurate timing of the eclipses of Jupiter's moons. The letter came to light only after Maxwell's death in 1879, aged 48 (see *Nature* **21**, 314–315; 1880). Had the data been available during Maxwell's life, there is every chance that he would have anticipated the Michelson–Morley experiment, and that he might have seen the need for special relativity two decades before Einstein. He was that kind of man.

Speculation of this kind belongs, of course, to the disreputable school of the history of what might have been. For Mahon, it is enough that Maxwell paved the way for the revolution to come. Maxwell may have been Mahon's hero for the past 40 years, but his book is not hagiography. General readers will glean from it not merely an absorbing account of Maxwell's life but also an explanation of why his work is at the foundation of the modern world. ■

John Maddox, now Emeritus Editor, was Editor of *Nature* for the periods 1966–73 and 1980–95.

Harnessing science for Hitler

Hitler's Scientists: Science, War and the Devil's Pact

by John Cornwell

Viking: 2003. 544 pp. £20, \$29.95

Kristie Macrakis

Adolf Hitler's ignorance about science and its utility during the early years of the Third Reich have become part of legends passed down by scientists and administrators for the next generation. The most-often repeated anecdotes involve Hitler's meetings with Max Planck, Nobel prizewinner and science administrator, and Albert Speer, Hitler's armaments minister. According to Planck, when he tried to persuade Hitler that doing away with Jewish scientists might be harmful, Hitler reportedly dismissed the idea saying: "So we'll do without science for a few years." Even during the Second World War, when leaders had begun to see the utility of science for the war effort, Speer related in his postwar memoirs that the concept of the atomic bomb "strained Hitler's intellectual capacity".

The title of this book implies that

Hitler had a stable of scientists who worked under his direction. This was not the case, but many scientists, physicians and engineers who worked in National Socialist Germany researched areas that supported National Socialist ideology. Others just happened to live and work there. To a certain extent, then, Cornwell distances himself from the Hitler leadership myth and acknowledges the existence of a "polycratic" regime run by competing Nazi institutions such as the army or civil service.

Cornwell, an author and journalist, is known for his best-selling book *Hitler's Pope*, which is based on archival research. But he admits that *Hitler's Scientists* is not based on original material. Instead, he has drawn together a history of German science in the first half of the twentieth century, concentrating on the Third Reich from a wide reading of secondary sources.

The time is ripe for this book. Over the past 30 years, historians have unearthed new material and provided fresh interpretations on topics such as Fritz Haber — chemical warfare pioneer, Nobel prizewinner and Jew — racial hygiene, medicine, physics, the German atomic bomb, rocket science and the leading German scientific institution, the Kaiser Wilhelm Society. But this is the first book to pull these themes together.

Cornwell has written an engaging synthesis of the original research: articulate and intelligent, with an eye for telling detail or anecdote. His lively account is also a damning indictment: many scientists come across as depraved, amoral nerds who were willing to serve any regime if they got paid for it. Cornwell unearthed a quotation from Wernher von Braun, designer of the German V-2 'flying bomb' who went on to direct NASA's Apollo programme, illustrating that "he did not care if he worked for Uncle Joe or Uncle Sam: 'All I really wanted was an uncle who was rich'."

Even though much of the book relies on previous work, Cornwell puts his own stamp on the literature and sometimes tells the story better than the specialists. There are also some refreshing new slants: most of the recent academic literature ignores decisively important areas for the war effort such as radar, submarines and codes.

Despite his familiarity with the recent literature, which usually provides a more nuanced account of science in the Third Reich than the contemporaneous reports and early postwar histories, Cornwell sometimes reverts back to the tone and interpretation of those earlier accounts.



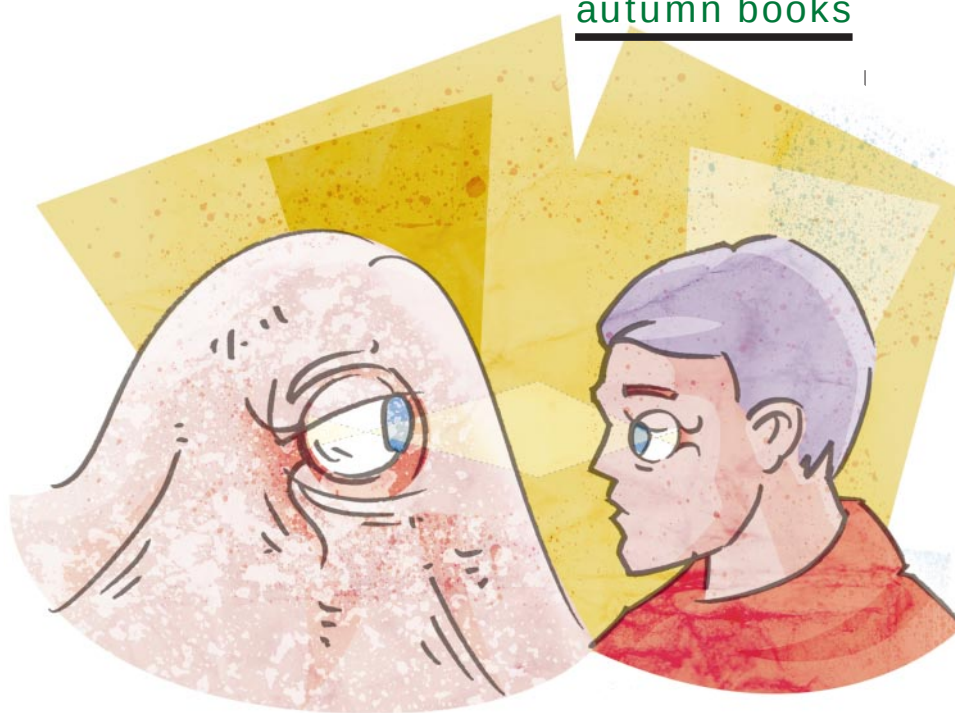
The author tasted the flavour of the world's attitude towards the Germans when he was a child growing up in England. He states that an aphorism that echoed through his boyhood and youth was: "The only good German is a dead German." Others saw something "congenitally malevolent about German people". Some of this anti-German attitude has remained in his psyche and surfaces from time to time in the book, even though he engages the newer interpretations. According to Cornwell, no scientist who worked in Nazi Germany is exempt from blame; even an academic scientist working on the purest research is guilty by association. Cornwell, however, wants all scientists to have a conscience, not just those who worked during the Nazi period.

Cornwell extends his period to reflect on the terrorist attacks of 11 September 2001 and technology, and adds material that seems over-detailed and tangential. But like other recent books on science under the Nazis, the real story begins a century earlier. The first quarter of the book describes Germany in the early twentieth century as a scientific mecca, with scientists such as Haber developing poison gas for the fatherland. The science of eugenics, later so important to the Nazis, was also born in this era. Others have written about how science "survived the swastika", but Cornwell emphasizes the strain of the Weimar period, when Germany was under the yoke of the Versailles Treaty and suffered because of the resulting economic problems. Science flourished there nevertheless.

The more sensational topics, such as medical experiments and slave labour, receive the most attention. There is one chapter on sciences that flourished under the Nazis, which includes a summary of Robert Proctor's book *The Nazi War on Cancer* (Princeton University Press, 1999). But Cornwell ignores the literature that demonstrates that basic biological research also survived and thrived; he thinks it just stagnated. He also fails to mention the anomaly that three Nobel prizes were awarded to scientists of the Kaiser Wilhelm Society in the 1930s and 1940s for work they did during the Nazi period: Richard Kuhn in 1938 for his work on carotenoids and vitamins; Adolf Butenandt in 1939 for his work on sex hormones; and Otto Hahn in 1944 for his 1938 discovery of nuclear fission. He does, however, add some original research and comments to the controversy surrounding Werner Heisenberg's visit to Niels Bohr in Copenhagen.

Even if *Hitler's Scientists* is not based on original research and not all scientists belonged to Hitler, it is a useful compilation for readers who would like just one volume on science under the Nazis.

Kristie Macrakis is in the Department of History and the Lyman Briggs School of Science, Michigan State University, East Lansing, Michigan 48825, USA.



The eyes have it

Life's Solution: Inevitable Humans in a Lonely Universe

by Simon Conway Morris

Cambridge University Press: 2003. 464 pp.

£18.95, \$30

Richard E. Lenski

The late Stephen Jay Gould argued that evolution is unpredictable. His thought-experiment of replaying life's tape, in *Wonderful Life* (W. W. Norton, 1989), led him to assert: "Alter any early event, ever so slightly and without apparent importance at the time, and evolution cascades into a radically different channel." The aim of *Life's Solution* is to overthrow that historically contingent view of life. What is at stake, according to its author Simon Conway Morris, is more than the statistical mechanics of evolution: at stake is how we understand our place in the Universe.

Life's Solution builds a forceful case for the predictability of evolutionary outcomes, not in terms of genetic details but rather their broad phenotypic manifestations. The case rests on a remarkable compilation of examples of convergent evolution, in which two or more lineages have independently evolved similar structures and functions. The examples range from the aerodynamics of hovering moths and hummingbirds to the use of silk by spiders and some insects to capture prey.

Convergence is widespread, despite the infinitude of genetic possibilities, because "the evolutionary routes are many, but the destinations are limited", as Conway Morris puts it. Certain destinations are

precluded by "the howling wildernesses of the maladaptive", where the vast majority of genotypes are non-viable and prevent further exploration by natural selection. Conway Morris is spectacularly successful at tracking down and organizing examples of convergent evolution, but he admits that work to place convergences "into any sort of quantitative framework is still in its infancy". In effect, he emphasizes the numerator (convergence) while skirting the denominator (all examples of evolution, both convergent and divergent).

Conway Morris is not content, however, to catalogue examples of convergent evolution. He wants to convince us that convergence implies the inevitability that some sentient human-like being will evolve on any life-bearing planet like Earth. Thus, he focuses his compilation on the attributes we associate with ourselves and the lineage that produced us. Fruiting bodies of slime moulds and myxobacteria show that multicellularity has evolved repeatedly. Warm-bloodedness evolved several times, as did live birth and even penile tumescence. Sensory organs exhibit numerous cases of convergence: the eyes have it, as seen in the camera-like eyes of vertebrates and octopuses, and the similar eyes of certain worms and jellyfish. So, too, mechanisms used by diverse organisms to smell, hear, echolocate, sense electrical fields and maintain balance are often convergent.

Complex social systems have evolved repeatedly, exemplified by termites, ants and mole-rats. Fungus-farming ants and tool use by crows remind us that even agriculture and invention have evolved elsewhere. Just as the availability of light to guide organisms has led repeatedly to the evolution of eyes,

the acquisition of complex information about an organism's environment implies selection to better store and use the information. The evolution of extra-large brains and corresponding intelligence in dolphins and primates illustrates that convergence. Conway Morris stops short of saying that another species has yet evolved the sophisticated language of humans, but suggests that "waiting in the wings of the theatre of consciousness are other minds stirring, poised on the threshold of articulation". These are provocative words and, whether or not one agrees with his conclusions, the examples are fascinating and indicate a prodigious knowledge of the scattered literature on convergent evolution.

Given this rampant convergence here on Earth, Conway Morris believes that "extra-terrestrials with nervous systems will hear, see, and smell in very much the same way as we do, and if that is so will also possibly have similar mental processes". So where are these ETs? Alas, he doubts that they exist. Conway Morris and Gould both think that we humans might be alone, but for different reasons. For Gould, "the awesome improbability of human evolution" derives from contingency in adaptive evolution. Conway Morris argues that if our planet were even slightly different from the way it actually is, then life might not have emerged. His argument is based on the difficulties of getting life started, on the failure of scientists to synthesize life from scratch, and on some unusual features of Earth and our Solar System. He even suggests that intelligence might never have evolved here had not a cataclysmic impact jettisoned the Moon into its orbit. This sounds rather like Gould's historical contingency, except that Conway Morris emphasizes physical events creating opportunities for life to emerge and adapt, whereas Gould emphasized the idiosyncratic nature of adaptation itself.

The tension between inevitability and loneliness leads Conway Morris towards a higher objective, which is to re-establish "notions of awe and wonder" in evolution and thus "allow a conversation with religious sensibilities". He dismisses Fred Hoyle's "strange ideas about the origins of biological complexity" but admits a grudging respect for Hoyle's remark that the Universe is a "set-up job". Conway Morris's metaphysical vision occasionally becomes overwrought, as when he says: "Not only is the Universe strangely fit to purpose, but so, too, as I have argued throughout this book, is life's ability to navigate its solutions." Whatever Conway Morris may think about the Universe and its predispositions, *Life's Solution* invokes the standard darwinian explanation of adaptation by natural selection for life's ability to navigate.

I recommend this book to anyone grappling with the meaning of evolution and

our place in the Universe, and to biologists interested in adaptation and constraints. I am obliged, however, to caution readers about the deprecating way in which Conway Morris sometimes refers to evolutionists whose views he opposes. He is especially dismissive of Gould, who died a year ago: readers interested in their conflict can read an exchange elsewhere (*Natural History* **107**, 48–55; 1998). Conway Morris's antagonism to Gould becomes more puzzling when one reads — in a chapter titled "Towards a theology of evolution?" — of his disdain for "ultra-Darwinists" and "genetic fundamentalism", as these were also frequent targets of Gould's pen. But while Gould argued for the separation of science and religion, Conway Morris is searching for common ground.

Conway Morris derides the "almost gleeful abasement of humans" by ultra-Darwinists, and claims that Darwin himself "retreated into a gloomy agnosticism". But the closing passage of *The Origin of Species* is far from gloomy: "There is grandeur in this view of life, with its several powers having been originally breathed into a few forms or into one; and that ... from so simple a beginning endless forms most beautiful and most wonderful have been, and are being, evolved." In the second edition Darwin inserted three words (italicized here): "... breathed by the Creator". In *Life's Solution*, Conway Morris has perhaps explained why, in his view of life, the second edition might be preferable to the first. ■

Richard E. Lenski is in the Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, Michigan 48824, USA.

Listen, learn and construct

Constructing a Language: A Usage-Based Theory of Language Acquisition
by Michael Tomasello
Harvard University Press: 2003. 388 pp.
\$45, £29.95
Eve V. Clark

In 1965, Noam Chomsky posited that linguistic theory should be able to account for how children acquire a first language. In so doing, he triggered debates that have lasted ever since. What is innate? What and how much language do children hear? Are children's errors corrected? Do children use two different mechanisms for learning grammar: one for regularities in syntax and morphology that can be described with rules, another for irregular forms?

But few linguists have spent much time looking at language acquisition itself, pre-

ferring to debate the logic of the enterprise within linguistic theory. They have tended to ignore findings from studies of acquisition that are inconsistent with their favourite theories. In short, they have not recognized that describing a language is not the same as describing the process by which people acquire it.

By contrast, in the past few decades, psycholinguists have documented many of the facts of acquisition. They have examined what speech children hear, studied the processes for learning complex systems, and identified factors that influence development. But they have also generally ignored changes in syntactic theories — descriptions of the rules that govern language — and the associated issue of just what is innate about language in humans.

Tomasello has added a new perspective to these debates from the psycholinguistic side, based on his work with primates as well as children. He has brought together a number of the topics that psycholinguists have worked on: language studied as a system for communication, the relationships between language, memory and attention, how inferences about meaning are made in context, choices of conceptual perspective — the decision to call a dog a "dog" rather than an "animal" — how common ground is built up in communication, and how a speaker's intentions are interpreted. He emphasizes that language is essentially social and that it relies not only on vocabulary and linguistic constructions, but also on non-linguistic elements such as gesture and gaze.

He starts from the premise that children acquire language by attending closely to the language they hear. To do that, they must analyse speakers' intentions and find any patterns in the language that speakers use. Tomasello argues that children acquire constructions in the same way as they do words: they have to learn both, and just as they slowly build up their vocabulary, they also slowly build up a repertoire of constructions. Words, in fact, are stepping-stones to constructions. For instance, children first use a verb like "want" only with "that" ("want that"), then with a following verb ("wanna go"), and only later still with a direct object and following verb ("want him [to] go out"). They build up larger constructions by combining smaller ones. Because Tomasello looks at speaker intentions as well as patterns of use, he integrates the cognitive and the social in language development from the start.

Researchers have always assumed that children acquire vocabulary by learning, but many have argued that learning alone can't explain children's acquisition of the regularities of language that can be described in rules for syntax and morphology. Acquisition of these, propose Steven Pinker and others, depends on innate language-specific



capacities. These capacities, argue researchers, are shared by all languages and include innately given word classes.

But in Tomasello's view, acquiring a first language entails mastering more than its grammar. It means learning to use the language to communicate, using the same resources that adult speakers do. The child's abstraction of grammatical rules, as sketched out in Chomsky's proposal, remains an important part of this task but, as Tomasello points out, it is unclear how quickly children identify such rules.

Tomasello presents a wealth of observation and argument in support of his approach. He appeals to recent linguistics research on constructions in syntax, and to psychological research on children's understanding of intentions and beliefs in others, on joint attention in communication, and on function-based distributional analyses and analogy in learning. He makes a compelling case for his view of acquisition as an alternative both to those linguistic accounts that have focused on grammar and on how much is innate, and to current 'connectionist' accounts by Jeffrey Elman and his colleagues that focus on the forms learnt but not on their meanings. He argues against the idea that there are different mechanisms for learning rule-based and irregular forms of language, and in favour of a single mechanism for

learning both words and constructions. And, like many biologists, he cautions against assuming innateness without examining the alternatives.

Tomasello should make us think more, and more carefully, about language in social as well as cognitive terms, and to consider the roles of attention, memory and learning in the process of acquisition. But he also leaves many questions unanswered. For example, what are the units of language being learnt? How should one define them — the notion of clause, for example? When do children learn to rely on conventions — that "dog" designates the category of dogs in English, but "chien" does so in French? Did languages evolve in the kinds of settings where adults and infants first establish joint attention? How do children learn the meanings of words and constructions? How is the choice of a conceptual perspective by the speaker — "that dog" or "that animal" — related to the build-up of common ground in conversation? How do children get rid of errors such as "comed" (for "came") or "me throw" (for "I want to throw it") in their speech? Are they attentive to corrections from adults, and if so, how? Are they, in fact, exposed to enough information about constructions to allow for learning?

These are all empirical questions that must be taken seriously. How children

acquire language can no longer be presented as a thought-experiment about the language that is the product of acquisition: it demands concrete data and theory about the process of acquisition. ■

Eve V. Clark is in the Department of Linguistics, Stanford University, Stanford, California 94305-2150, USA.

Changing the world

Niche Construction: The Neglected Process in Evolution
by F. John Odling-Smee, Kevin N. Laland & Marcus W. Feldman
Princeton University Press: 2003. 486 pp.
\$75, £49.95 (hbk); \$39.50, £26.95 (pbk)
Laurent Keller

"Niche construction changes our conception of the evolutionary process ... and should be regarded, after natural selection, as a second major participant in evolution." So argue the authors of this book as early as page 2. But what exactly is niche construction? And if it's really so important, how could evolutionary biologists have neglected it for so long?

The concept of niche construction is relatively simple. All living creatures, through both their metabolism and their behaviour, actively change and control the world in which they live. Organisms choose habitats and resources; they construct nests, holes, burrows, webs or pupal cases; and they modify the chemical environment in which they live. These alterations, which occur at scales ranging from the extremely local to the global, inevitably modify some of the selection pressures acting on the organisms. And it is precisely this — the effects of an organism on its own environment — that the authors believe to be the important component that has been neglected by the conventional theory of evolution.

The concept that organisms bring about important changes in their environment that may in turn affect their fitness is not completely new, however. Charles Darwin had already made several potent observations to this effect. For example, in *On the Origin of Species* he states: "When a species, owing to highly favourable circumstances, increases inordinately in numbers in a small tract, epidemics — at least this seems generally to occur with our game animals — often ensue: and here we have a limiting check independent of the struggle for life." This observation is strongly reminiscent of one in *Niche Construction* in which the authors discuss how large-scale human aggregation resulting from the construction of villages, towns and cities may create new health hazards such as the spread of epidemics.

In the 1980s, several scientists pointed



out that organisms not only adapt to their environment but construct it as well. In *The Extended Phenotype* (Freeman, 1982), Richard Dawkins argues that genes can express themselves outside the bodies of the organisms that carry them, an example being a beaver's dam. Richard Lewontin, in his contribution to *Evolution from Molecules to Men* (Cambridge University Press, 1983), also realized that the histories of organisms and the environment are a function of each other. He therefore suggested that what is actually happening in nature could be represented by a pair of coupled differential equations.

Niche Construction goes one step further by providing an exhaustive list of the possible effects of organisms on their environment, from the alteration of non-living environments to various sorts of maternal effects induced by differences in the levels of yolk, hormones and messenger RNA in the cytoplasm. In humans, cultural processes constitute an additional kind of non-genetic information that is transmitted from one generation to the next. The authors correctly argue that niche construction may result in evolutionary feedback, with the evolutionary trajectories of organisms being influenced by the changes that they induce on their environment. Feedback also occurs across generations, with individuals being influenced by the environmental modifications provoked by their ancestors.

To make their point, the authors list numerous examples. One of the most

striking is the evolution of lactose tolerance. The domestication of cattle brought milk and dairy products into the diet of some human populations for enough generations to promote genes that confer greater lactose tolerance. Another example of a culturally induced genetic signature relates to the influence of agricultural niche construction on natural selection. The Kwa-speaking yam cultivators of West Africa made clearings in tropical rainforests, increasing the amount of standing water. This provided superior breeding grounds for malaria-carrying mosquitoes, which in turn increased the prevalence of malaria. At the same time, the frequency of the haemoglobin allele responsible for sickle-cell anaemia increased, as people who have one copy of this allele have some resistance to malaria. These examples demonstrate how cultural processes are not just a product of human genetic evolution but also a cause of it.

So does this mean that we need a new, extended theory of evolution as advocated by *Niche Construction*? I believe that the answer is no. First of all, the examples of niche construction given in the book can be explained by conventional evolutionary theory once all of the relevant environmental factors are taken into account. Several elegant models—including some developed by the authors themselves—have already been proposed to deal with feedback effects between organisms and their environment. At the same time, it is not possible to develop

an extended theory of evolution that can encompass all of the different types of niche-construction effects. Interactions between organisms and their environment are, by their very nature, complex and result in an intricate web of interacting effects. No heuristic theory could integrate all of these effects without becoming intractable, so we need to devise specific models for each situation. Indeed, evolutionary biologists are already doing exactly this.

To summarize, *Niche Construction* does an excellent job of detailing the many ways in which organisms modify their environment and hence the selective forces acting on them. But it is unfortunate that the authors attempt to oversell the significance of niche construction. By advocating a grand, extended evolutionary theory, they distract readers from the more important message of the book, which is that the influence of organisms on their environment can have far-reaching consequences. Indeed, humans are currently working hard at building a vivid example of this. By allowing the massive-scale niche destruction that is currently under way, we are not only compromising the environment, but also affecting the prospects and evolutionary trajectories of our children and many later generations. ■

Laurent Keller is in the Department of Ecology and Evolution, Biology Building, University of Lausanne, 1015 Lausanne, Switzerland.

..... An enigma for all seasons

Prof: The Life of Frederick Lindemann

by Adrian Fort

Jonathan Cape: 2003. 374 pp.

£18.99, Can\$48.50

Stuart Young

According to Mark Twain, "Biographies are but the clothes and buttons of the man — the biography of the man himself cannot be written." Yet with so puzzling a subject as Frederick Lindemann, attempting to portray "the man himself" is difficult to resist. In the world of science he was exceptional, and in the world at war his power was as great as that of any scientist in history. But he could also be reticent and reclusive. Although former associates remember him well, to many others Lindemann remains, as was said of him recently, "the most important man you have never heard of".

Lindemann's cherished homeland was England, but his scientific education and early research took place in Germany, where he worked successfully on thermodynamic problems with Walther Nernst in Berlin. In 1911, aged just 25, Lindemann took part in

conference discussions with all those scientists whose names were so prominent in the evolution of modern physics, including H. A. Lorentz, Max Planck, Albert Einstein, Ernest Rutherford and Marie Curie. Back in England during the First World War, he turned his experimental skill to warplanes, which he learned to fly to prove his theoretical method of recovering from an aircraft spin. And then in 1919, as professor of physics at Oxford University, he began the major task of recreating the Clarendon Laboratory's teaching and research facilities.

It is part of the puzzle that during his lifetime Lindemann aroused antipathy, if not worse, in some quarters, yet great affection and loyalty in others. His abiding and closest friendship was with Winston Churchill, who constantly relied on him before, during and after the Second World War for advice, not only on strictly technical matters, but also on a wide range of topics that his scientific mind could clarify. In the war years, Churchill required, and Lindemann supplied, concise summaries of the situation at the startling average rate of about one a day. They covered bombing strategy, shipping capacity, aircraft defence, strategic supplies, explosive efficiency, German 'flying bombs' and many other topics.

The persistence with which Lindemann's group of experts probed for the data, and his obstinacy and scorn when challenged, brought him into conflict with many at the time. After his death in 1957 there were some who vilified him as a malign hindrance to the Allied war effort, whereas others hailed him as one of the main contributors to its success. And throughout, his

idiosyncrasies were thoroughly documented, as if they added significance, rather than mere colour, to the events in his life.

Until now, the foremost objective account of Lindemann's life has been *The Prof in Two Worlds* by the Earl of Birkenhead (Collins, 1961). This was prepared as the official biography soon after Lindemann's death in 1957, although in the event it was pipped at the post by *The Prof: A Personal Memoir of Lord Cherwell* by R. F. Harrod (Macmillan, 1959).

The new biography by Adrian Fort is timely. He has searched widely to uncover every available scrap of relevant information about his subject. It is clear that private papers and official documents were opened to him, shedding new and brighter light on some parts of Lindemann's influential life. But the key word is 'available', and the story told by Fort begs the question: how much more may still be hidden in secret papers? To take one example, did Lindemann have any influence in the vital wartime business of coding and decrypting?

Prof contains extracts from some of Lindemann's wartime minutes to Churchill. These provide an interesting complement to Churchill's own published account of

the events, in which the directives all flowed in the opposite direction.

This is a well-rounded, very readable and reasonably illustrated life story of a remarkable man who had a profound influence on academic and political affairs. It shows how between the wars he was almost solely responsible for the rebirth of physics at Oxford University, and how the achievements there in low-temperature research were the consequence of his personal and timely protection of some prominent scientists — including Francis Simon, Kurt Mendelssohn and Nicholas Kurti — who were under threat from Nazi persecution.

With Churchill, Lindemann managed at the eleventh hour to reverse Britain's unpreparedness for air warfare, and he promoted decisive wartime advances in radio-wave air defence. Significant parts of the ensuing war strategy, such as the allocation of bomber aircraft, for which he has been heavily criticized, were a consequence of Lindemann's advice. In the 1950s he won a lone and strenuous battle to create Britain's atomic-energy enterprise. And during this period he fostered a practical demonstration of one of his interests — a laboratory at Oxford dedicated to the application of scientific methods in art history and archaeology.

Mention of Lindemann's prominent and public activities bypasses the account of his personal research achievements. Fort sketches these with care, although readers should not expect to obtain from him more than a simple, but well-referenced, summary. The large range of Lindemann's experimental and theoretical interests is, in fact, covered by more than 60 published papers.

His letters to his father reveal the softer side of Lindemann's nature and may encourage students of experimental physics. For example, there is one, written from Nernst's laboratory in Berlin, that reveals all the excitement he felt, yet contained by commendable caution, when on the brink of what he thought might be a new discovery. Other letters relate to his progress in the company of all the greatest names of the revolution in physics in the early twentieth century. And for those who may think that top-flight physics must be all-absorbing, it is worth more than a glance to see how Lindemann's student days were spent.

I enjoyed reading this book, which added considerably to my understanding of Lindemann — not only of his great impact, but now also something more of that very intriguing man himself. ■

Stuart Young is at Box 4, Noordhoek 7985, South Africa.



A feather in the cap for bird breeders

The Red Canary: The Story of the First Genetically Engineered Animal

by Tim Birkhead

Weidenfeld & Nicolson: 2003. 284 pp. £16.99

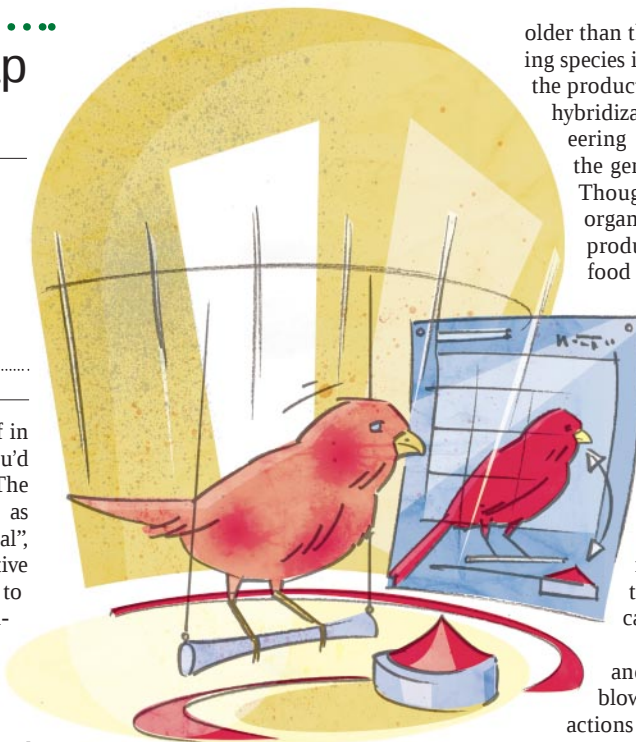
(Published in the United States as A Brand-New Bird. Basic Books, \$26)

Jerry A. Coyne

Plucking *The Red Canary* off the shelf in a bookshop, you'd assume that you'd found a book about biotechnology. The subtitle touts the eponymous bird as "the first genetically engineered animal", and the dust jacket promises a narrative of "huge contemporary relevance" to "the exciting but controversial developments in genetic manipulation". But it's not about biotechnology at all. It is in fact a book about animal breeding of the traditional sort — there's not a restriction enzyme, plasmid vector or polymerase chain reaction (PCR) machine in sight. Presumably the publicists' attempt to drum up interest in the book explains the strange disconnect between its outside and its inside: they've given *The Red Canary* showy but artificial plumage.

A more appropriate but less attractive title would be *A History of Canary Breeding*. For the story of the red canary occupies only half of an account that begins in the fifteenth century when wild canaries were first snatched from — where else? — the Canary Islands. At first the preserve of aristocracy, pet canaries were soon a staple of European life, beloved for their colour and beguiling song. Canary breeding became a cottage industry in Germany, which pumped out thousands of 'Harzer rollers' (named for their rumbling call) for the European market. Birkhead notes that in 1742 there were roughly 200,000 canaries in London alone.

It's an absorbing if not riveting tale, and Birkhead brings illumination and ornithological insight to a dark and neglected corner of animal breeding. We learn, for example, that for practical reasons, breeders and bird-catchers pioneered the study of bird migration and behaviour, inadvertently fledging a new science — ornithology — in the process. And there are plenty of interesting tidbits, such as the peculiar origin of a musical instrument: the recorder was invented to train songbirds. Nor does Birkhead neglect the role for which canaries are proverbial: they were used as an early-warning system in coalmines because their rapid respiration made them especially susceptible to toxic gas, while their song brightened the gloom underground.



The core of the book is the unlikely collaboration between two Germans: Hans Dunker, an academic ornithologist, and Karl Reich, a shopkeeper and fanatical canary hobbyist. With Reich's expertise, Dunker worked out the genetics of canary song and colour, gaining modest fame in the process. Around 1925, they set out to create the avian equivalent of the blue rose: a red canary. They 'genetically engineered' the bird by crossing it with a close relative, the South American red siskin, and then repeatedly crossing the reddest hybrids back to canaries. But despite years of effort, their project failed. Their canaries became no redder than an unattractive, blotchy orange. When National Socialism intervened, the enterprise was forgotten as Dunker abandoned canaries for a new role as an enthusiastic propagandist for Nazi eugenics.

The fabled red bird was not produced until 1964, and then only through chicanery. Jack Swift, a British breeder, discovered that dark orange birds would become red prize-winners when fed carotenoids — the same substance that turns flamingos pink. And so the Holy Grail of canarydom was reached by manipulating both genetics and diet.

To convince us of his tale's significance, Birkhead plays up two themes: genetic engineering and the connection between genes and environment — nature and nurture. But the effect is the opposite of what Birkhead intended: his digressions remind us that, for all of their charm, canaries have little to offer on either theme.

Birkhead shows that genetic engineering — at least, when defined as the transfer of genes between species — is not new, but has been around for at least 75 years. In fact, it is

older than that: orchid breeders were crossing species in the 1850s, and bread wheat is the product of several ancient interspecific hybridizations. But the genetic engineering of Dunker and Reich is not the genetic engineering of Monsanto. Though they both yield transgenic organisms, the issues surrounding the production of genetically modified food and of red canaries are worlds

apart. Canaries are irrelevant even when we ignore the technical issues and envisage the worst-case scenarios — the escape of bacterial-toxin genes from genetically engineered corn is far more worrisome than the escape of colour genes from canaries. To imply that Reich and Dunker's experiments offer insight into modern transgenics is to turn the red canary into a red herring.

The interaction between genes and environment is also overblown. It is true that such interactions are important in many areas of genetics, as shown in Matt Ridley's recent book *Nature Via Nurture* (Fourth Estate, 2003). And it is true that it takes carotene as well as genes to make a red canary, just as it takes both pulchritude and peroxide to make a Marilyn Monroe. But Birkhead goes too far when claiming that the gene-environment interaction is "one of the most fundamental phenomena in biology because it recognizes that the nature versus nurture debate is erroneous and meaningless". The debate is not meaningless, because specific environments are not critical for much of animal breeding. A fertilized dachshund egg becomes an adult dachshund regardless of where gestation occurs or what one feeds the puppy.

As a biologist, my enthusiasm for *The Red Canary* — admittedly a book for the layman — was further eroded by scientific gaffes that too often mar the discussion of genetics. Birkhead, for example, explains that species hybrids are especially vigorous because only one gene of each pair is expressed, and that the "genes that are turned on are those that best complement the rest of that individual's genetic make-up". But even if hybrids were generally more vigorous than pure species, which they aren't, this explanation is dead wrong, literally: hybrids expressing only half their alleles would not survive.

While *The Red Canary* tells a good tale, it suffers from its need to place the story in a broader context. It provides little insight into controversies about genetic engineering, nor does it illuminate the nature-nurture debate. Burdened with these expectations, *The Red Canary* flutters gamely but fails to soar. ■

Jerry A. Coyne is in the Department of Ecology and Evolution, University of Chicago, 1101 East 57 Street, Chicago, Illinois 60637, USA.

The observer

Summarize yourself in the form of a title of a paper in Nature.

Vera Rubin studies how stars orbit their galaxies and how galaxies move in the Universe.

What was your first experiment as a child?

Two chunks of ice on a sunny step with a black cloth over one and a white cloth over the other, to watch which one melted first. I was impressed.

Who has been the most important mentor in your career?

My husband.

Whose graduate student would you most like to have been?

I wrote my PhD thesis under George Gamow, and it's hard to better that. Publicity arising from my Cornell MA thesis prompted Gamow, a physics professor at George Washington University, to call me to discuss the space distribution of galaxies. His piercing questions prompted me to arrange to write my PhD thesis under Gamow's direction.

What book has been most influential in your scientific career?

The Hubble Atlas of Galaxies by Allan Sandage (1961), because the stunning images added to the puzzles that galaxies provide.

What gives you the most job satisfaction now? What are your major frustrations?

Satisfaction: observing, and continuing to learn new things about galaxies. Frustration: too much e-mail.

What's your favourite conference destination?

Florence, Italy. What other venue combines a science museum containing Galileo's telescope and old globes, with wonderful Renaissance painting, sculptures and great food?

You have the audience in your hands, but some smart-alec asks you the killer question you have no idea how to answer. What's your response?

"I have no idea how to answer."

What books are currently on your bedside table?

Marie Curie by Susan Quinn (preparing for the 100th anniversary of Curie's 1903 Nobel Prize in Physics); *Rosalind Franklin* by Brenda Maddox; *Hypatia of Alexandria* by Maria Dzielska; and *The Center of Things* by Jenny McPhee.

You are on a plane behind two students obviously going to the same conference, who start to talk about your work. What do you do?

Tap them on the shoulder and tell them I'm listening.

What music heads the playlist in your car or lab? Anything by Stephen Sondheim.

Assuming the dead can be raised and/or time travel exists, who from the world outside science would you most like to have dinner with?

Benjamin Franklin.

Where and when would you most like to have lived or worked?

It couldn't be any better than here and now.

What one thing would you rescue from your burning laboratory?

Years ago it would have been photographic plates. Now, probably recent telescope logs, although most comments are recorded via computer.

What do you do to relax?

Garden.

What does your garden grow? Have you any gardening advice for busy scientists?

Usual things, plus three witch-hazel trees: one (native Virginia) blooms in autumn, the other two in late winter. Quite magical. Grow lots of big green leafy things, so the garden mostly looks OK.

What's the best piece of advice you've received?

From Harold Urey: "I find that if I forget about my mistakes, everyone else forgets about them too." The worst piece (from another scientist): "It's not worth jeopardizing your career by your activities on behalf of women."

What would you have become, if not a scientist?

Miserable and unhappy.

What's your motto?

Don't give up; make it happen; ignore bad advice.

What single discovery, invention or innovation would most improve your life?

A secret day for me every week, that was not available to others.

You've just been told (in confidence) that the world will end tomorrow. What do you do next?

Suspect that it was not true, but call my husband.

Why is physics so hard?

Because of the usually awful way it is taught.

Which actor would best portray you in a film of your life story?

I would really like an actress who has studied physics or astronomy if there is such, but I'd settle for Dame Judi Dench. (I aim high.)



Vera Rubin

Vera Rubin, senior fellow in the Department of Terrestrial Magnetism, Carnegie Institution of Washington, is wife, parent and astronomer, in that order. She and husband Bob have four PhD offspring: two geologists, an astronomer and a mathematician.

Which field in science (apart from your own) deserves more funding, and why?

That's too hard — what does 'deserve' mean?

Oh, go on, give it a try.

Maybe the science of teaching and learning. Why is it so hard in the United States to raise the reading and maths scores of young students?

What overlooked or underrated discoveries really changed the science in which you work?

Women assistants (discovered by Pickering at Harvard) and money (discovered by Hale, who three times was responsible for the largest telescope in the world).

What's the most interesting thing in your fridge?

Chocolate. More chocolate.

Name one extravagance you can now get away with because of your eminence.

I can get a seat at a crowded seminar, but maybe that's because of my age.

Due to an unexpected change in Earth's rotation, days now last 25 hours. How would you spend the extra hour?

Worrying about the change in the length of the day.

The host at a dinner party has placed you next to God. What would you talk about?

The origin of the Universe, and why He/She did it that way. Also dark matter.

If you were reborn as a comic-book superhero, what would be your superhuman power?

Eyes to see to the origin of the Universe! ■

Six is seventh

Jane Grimwood and Jeremy Schmutz

The finished sequence of human chromosome 6 reveals an abundance of biological information previously buried within the draft of the human genome, and illustrates the increasing power of comparative genomics.

Monday 14 April 2003 was a momentous day. It was the culmination of a project that had consumed the professional and personal lives of an international group of researchers for as far back as most could remember. The Human Genome Project had surpassed its goals with the early completion of the 'finished' human genome: 2,825 million base pairs of DNA sequence, finished to exceptionally high quality and accuracy. The human genome is divided into 24 chromosomes (Fig. 1), each unique in its composition and contribution to the biological basis of being human. With the sequencing phase complete, each of the chromosomes is now being annotated, a process that involves an in-depth manual review and curation of the gene content. Six of the chromosomes (22, 21, 20, 14, Y and 7) have already been fully analysed and the results published. On page 805 of this issue¹, Mungall *et al.* describe the DNA sequence and analysis of the seventh completed and fully analysed chromosome, human chromosome 6.

This publication marks the end of a mapping, sequencing and finishing endeavour that began systematically in September 1996 at the Sanger Institute, Cambridge, UK. The finishing phase involves turning a rough draft sequence into a highly accurate finished sequence, with no sequence gaps and a defined error rate. Chromosome 6 weighs in at 166,880,988 base pairs, or more than 166 megabases, making it the largest human chromosome to be completely analysed and published so far. The finished sequence, comprising a little less than 6% of the total human genome, is in nine pieces, each divided by a small gap, or missing segment. All but two of these missing segments are located in regions of highly repetitive DNA near the centre or at the very tips of the chromosome arms. Mungall *et al.* estimate that they have captured 99.5% of the chromosome's euchromatin, the portion of the chromosome that excludes these repetitive areas.

What have we learned about the landscape of chromosome 6? Mungall *et al.* describe 2,190 gene structures, with evidence that 1,557 of them are functional genes and 633 are pseudogenes (inactive genes). Of the 1,557 that are predicted to be active or functional, only 772 have been previously described — a proportion



Figure 1 The big picture — the paired chromosome complement of a normal male, with 22 non-sex-determining chromosomes and an X and a Y chromosome.

comparable to that of the other published human chromosomes.

Many of the statistics for chromosome 6 are typical of the genome as a whole². For example, the average gene density per megabase is 9.2 (the genome average is about 10). And the percentage of repeated sequence is 43.95% (genome average 44.8%). Other attributes of chromosome 6 make it stand out from the crowd, however. They include the largest known cluster (157 genes) of transfer RNAs, which are involved in the translation of DNA sequence to amino acids, and the most polymorphic gene yet discovered — *HLA-B*, which shows unparalleled variation between individuals.

Chromosome 6 is best known as the home of the major histocompatibility complex (MHC), a region of 3.6 megabases that has an essential role in immune systems. Because of its importance and association with many common diseases, the complete DNA sequence of this region was published back in 1999 (ref. 3). But the MHC region should not overshadow the medical importance of several other genes on the chromosome. Examples are *HFE*, mutations in which cause hereditary haemochromatosis, a condition that affects 1 in 400 people and results in multi-organ dysfunction caused by increased iron deposition⁴; *EPM2A*,

mutations in which cause a disease known as Lafora's myoclonus epilepsy; and *PARK2*, point mutations or deletions in which are responsible for a juvenile-onset form of Parkinson's disease. Gene abnormalities on chromosome 6 are also implicated — although not yet confirmed — as a contributory cause of several genetically complex diseases, including schizophrenia, epilepsy, cancer and heart disease.

Adding the base pairs of chromosome 6 to those of the six others that have already been finished gives a total of 500 megabases of fully analysed human sequence. This is only 17% of the 2.8 billion base pairs of the total sequence. What is the status of the remaining 17 chromosomes that have yet to be completed? One severely underestimated part of the process is the difficulty of fully annotating a finished chromosome⁵. Various groups are still striving to deliver the most complete snapshot of the chromosomes within the limits of current gene-prediction science. In addition, although the original goals of the Human Genome Project were exceeded in finishing 99% of the sequence, work is continuing on the 'exceptional duplicated regions'⁶. These are chromosome segments that have been recently duplicated and are believed to be the birthplace of new genes, but classical chromosome mapping



100 YEARS AGO

While walking along the main road on the outskirts of Bordighera yesterday morning, I noticed a strange-looking insect moving across it in a peculiar way. On getting nearer, I saw that what had attracted my notice was a black ant — about an inch long with brown wings — dragging a cricket bigger than itself... A low wall ran alongside the road, and when the ant got within six feet of it a common brown lizard appeared on the top of the wall and evidently soon caught sight of the ant, for it ran quickly down the wall and to within two feet of it, when it crouched for a second or two like a cat ready to spring, and then charged the ant, apparently butting the cricket free with its head. Before the ant could regain its hold the lizard seized the cricket in its mouth, and darted up the wall in the direction from which it originally appeared on the scene, leaving the ant running round and round, moving its wings in an agitated manner, vainly searching for its lost prey.

From *Nature* 22 October 1903.

50 YEARS AGO

Montegazza observed the survival of human spermatozoa after exposure to a temperature of -15°C . He speculated that in the future, frozen semen might be used in animal husbandry and even proposed that a man dying on the battlefield might, by his wife, beget a legitimate child after his own death... Work in our laboratory, recently, indicated that treatment with 10 per cent glycerol prior to freezing with 'dry ice' produced an average 67 per cent survival of human spermatozoa obtained from five young healthy men... Three recipients have been successfully inseminated with frozen semen: one woman, nulliparous, has missed six menstrual periods, has cervical softening and uterine enlargement, and has developed other signs of pregnancy. The second woman, primiparous, has subjective signs of pregnancy, and one week after her first missed menses the Aschheim-Zondek test was suggestive; following an additional seven days of observation the test became positive. Five menses have not occurred. An X-ray film shows a normal developing foetal skeleton. The third recipient, nulliparous, has missed three menses and the Aschheim-Zondek test is positive.

From *Nature* 24 October 1953.

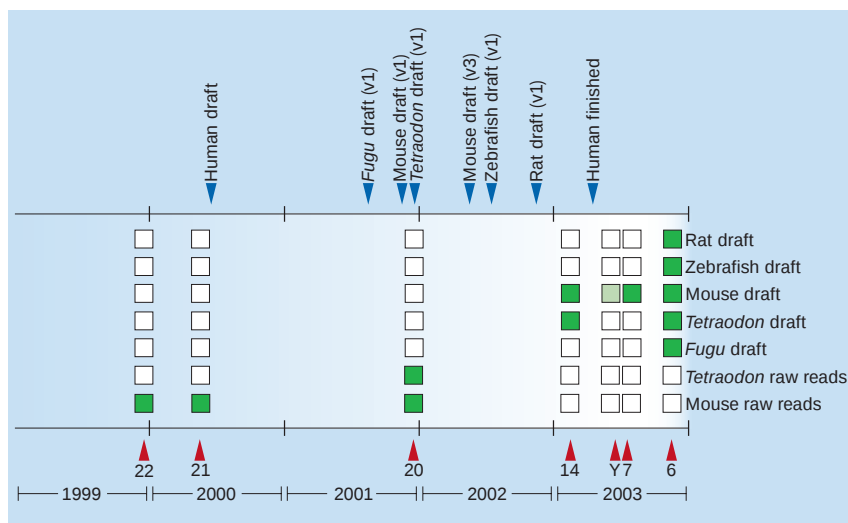


Figure 2 Comparative benefits. The increasing availability of draft genome sequences for other vertebrates is allowing better predictions of which stretches of human DNA constitute genes. This timeline shows the publication dates of the finished human chromosomes 22, 21, 20, 14, Y, 7 and 6. The green boxes indicate the cross-species comparisons that were included in the papers concerned; *Fugu* and *Tetraodon* are species of pufferfish that were chosen for sequencing because of their compact genomes. Sequencing proceeds from 'raw reads', to drafts of various versions, to finished; the light green box indicates the — obviously — limited comparison possible of the human Y chromosome with the draft sequence of a female mouse.

techniques have proved ineffective in teasing them apart.

However, full analyses of the remaining 17 chromosomes should be well worth the wait. As their description of chromosome 6 shows, Mungall *et al.*¹ have been the first to be able to fully apply the cross-species comparative power that is now on offer as the DNA sequences of other organisms become available (Fig. 2). Assembled draft genomes of the mouse, rat, *Tetraodon* (spotted green pufferfish), *Fugu* (tiger pufferfish) and zebrafish are in hand, and can now be used to search for regions of similarity or conservation between different organisms. This comparative approach allows refined predictions of which stretches of DNA are actually genes, and a more sophisticated interpretation of the underlying genomic data. The power of comparative genomics will grow as the genome sequences of the chicken, chimpanzee, frog, dog and cow, already in the production queue, become available.

Several other large-scale projects also have the aim of understanding how life functions at the biomolecular level. The newest of them, ENCODE (for 'Encyclopedia of DNA Elements')⁷, involves a microscopic examination of 30 million bases of the human genome (about 1%). The aim is to identify all of the functional elements and to create a type of *Peterson Field Guide* to DNA. This electronic book will contain descriptions, examples and identifying characteristics of each type of known DNA component; these can then be used to scan the rest of the genome and other genomes for similar pieces of DNA sequence that might have a

similar function. The ENCODE project encompasses both protein-coding elements and non-coding functional elements that do not themselves code for protein but instead regulate the production or expression of proteins from nearby genes.

Another project, MGC (the 'Mammalian Gene Collection')⁸, is attempting to amass at least one full-length messenger RNA transcript for each gene in the human and mouse genomes. Zebrafish and frog mRNAs are also now being included in this collection. These transcript sequences are vital for the continuing annotation of the human genome, in that they provide a complete picture of the entire gene after it has been spliced from the genome. Moreover, the mRNA clones can be introduced into a foreign host cell and made to produce a specific protein. These proteins can then be tested for functional activity. Similar projects are continuing worldwide, providing resources from a variety of organisms. All will help in interpreting the mysteries still buried in the DNA sequence of the human genome. ■

Jane Grimwood and Jeremy Schmutz are at the Stanford Human Genome Center, 975 California Avenue, Palo Alto, California 94304, USA.
e-mails: jane@shgc.stanford.edu
jeremy@shgc.stanford.edu

1. Mungall A. J. *et al.* *Nature* **425**, 805–811 (2003).
2. The International Human Genome Sequencing Consortium *Nature* **409**, 860–921 (2001).
3. The MHC Sequencing Consortium *Nature* **401**, 921–923 (1999).
4. Feder, J. N. *et al.* *Nature Genet.* **13**, 399–408 (2001).
5. Ashurst, J. L. & Collins, J. E. *Annu. Rev. Genomics Hum. Genet.* **4**, 69–88 (2003).
6. Eichler, E. E. *Genome Res.* **11**, 653–656 (2001).
7. www.genome.gov/10005107
8. <http://mgc.nci.nih.gov>

Applied physics

To catch a photon

Daniel E. Prober

Astronomers crave a detector sensitive enough to detect a single photon and determine its energy. A new single-pixel device can do this, and could also be built up into a large array suitable for a telescope.

Vision is the richest of the human senses, and detectors of light have long featured in science and technology. In fields as diverse as telecommunications, medicine and astronomy, there is demand for exquisitely sensitive detectors of the quantum of light, the photon. But as yet there is no single commercial device that can simultaneously detect individual visible photons and record the colour (or energy) and arrival time of each photon. If such a detector existed, it should also produce images with many picture elements (pixels) — and be affordable. Day and colleagues¹ now propose a device, based on the principle of kinetic inductance, that could function as an individual pixel in a photon detector (page 817 of this issue). Importantly, they show that the device has the necessary properties to allow the incorporation of

many such pixels into the 'ultimate' photon detector, one that could also be read out with practical, available electronics.

The search for the ultimate photon detector has been driven by astronomers, who are often faced with a limited number of photons to measure, emitted from some object in our Galaxy or beyond, and limited time in which to measure them. At present, astronomy is well served by the charge-coupled detector (CCD), familiar to many people as the CCD sensor in their digital camera or camcorder. Megapixel CCDs are now common. But this detector cannot resolve individual photons, because the noise occurring randomly in its readout electronics is too large to allow it. Moreover, the physics of the detector precludes the measurement of photon colour, unless colour filters are used. These reduce efficiency, but without such

filters a CCD would see only shades of grey.

To record single photons cleanly and to discern their energy and arrival time requires a detector that operates at low temperatures. This gets rid of the thermal agitation in the device that would disrupt a single-photon signal, and also means that materials and techniques can be used that are fundamentally different from those employed in the CCD. The first advance in cryogenic detector technology was the silicon 'microbolometer'², in which a small piece of silicon is held at 0.05 K but heats up when a photon is absorbed. The subsequent rise and fall of the temperature is recorded, to determine the photon energy. These detectors are employed by astronomers for rocket-based observations of photons at X-ray wavelengths, with an energy 100 to 1,000 times that of visible photons. But their sensitivity is not sufficient to record visible photons, with an energy of 1.5 to 3 electronvolts.

Over the past decade, new devices have been developed that are more sensitive and can record visible photons^{3–9}. These detectors fall into two classes. The first is the bolometer, which uses the same heating effect as mentioned above to determine photon energy. The thermometer inside such a device, recording the temperature change, is a strip

Genetics

Secrets of a porkier porker

Gregor Mendel's remarkable studies of peas revealed a straightforward pattern of genetic inheritance. But since then we've learned that matters are not always so simple: many identifiable traits result from complex interactions between multiple genes and environmental factors. Current work in genetics aims to quantify the contributions of specific variations in each gene to the mix. But pinpointing the right spot in an animal's huge genome can be a nightmare, even in organisms that have had their genome completely sequenced.

The humble pig, meanwhile, has only recently received its own genetic map — a first-generation sketch of what its genome sequence might look like. Imagine, then, how difficult it must have been to find one specific nucleotide that controls 15–30% of the variation in muscle mass seen between pigs. But that's just what Anne-Sophie Van Laere *et al.* report elsewhere in this issue (*Nature* 425, 832–836; 2003).

The authors find that a change from a guanine nucleotide to an adenine at a specific point in the



IGF2 gene can add 3–4% more meat to a pig. The nucleotide change disrupts the regulation of *IGF2*, increasing its expression threefold in muscle but not at all in liver, another main organ of expression. As the IGF-II protein stimulates muscle growth, the result is a pig with more muscle and less fat — a boon for meat production. And, as the change occurs in part

of the gene that does not actually code for protein, it suggests that researchers should not study just the protein-coding bits when looking for important genetic differences between individuals.

Of course, farmers don't need to know the exact mechanism involved; they spotted this physical trait years ago and selected for it in breeding schemes. Thus, Van Laere

et al. show that the muscle-favouring alteration has swept through commercial pig populations, but is not present in Asian or European wild boars tested. Now that the pig is lining up with other farm animals such as cows and chickens to have its genome sequenced, we should be able to find other genetic contributions to bulked-up pigs.

Chris Gunter

H. SCHWIND/OKAPIA/SPL

of partly superconducting metal^{3,6}, the readout amplifier is also superconducting. As superconductivity — the flow of current without resistance — is a low-temperature property, the entire structure of the device is a cryogenic environment (which is easier to engineer than a device with only some superconducting components).

The second class of detector makes use of the electron excitations created in solids by the energy of an incoming photon⁵. Numerous observations have been made with such detectors developed at the European Space Agency^{7–9}, including observations of the Crab nebula and short-period binary systems, known as polars⁹. The readout electronics of these detectors, however, operates at room temperature, with the consequence that the number of channels that can be read out is limited. So far, the number of pixels in an image has been limited to 36. Clever device design might mean that this number of pixels can be expanded by a factor of maybe 10 to 20, but the engineering is proving difficult.

Day *et al.*¹ now present a new approach to the problem of photon detection. Their device contains a superconducting film held at low temperature. Inside the film, resistanceless current flows in the form of electron pairs. If a photon hits the device, it can break up some of these pairs, with the result that the supercurrent becomes more 'sluggish'¹⁰. That increase in sluggishness can be detected in the surrounding microwave circuitry. The nature of the measurement is akin to watching an object on a spring, vibrating at the natural oscillation rate: the rate of the spring's vibration reveals the mass ('sluggishness') of the object.

The authors also show that cryogenic electronics systems already available are adequate for the readout of the signal from their device. Device and readout together make the desired single-photon detector. Moreover, the method of readout used can be rather easily generalized to accommodate hundreds, and perhaps thousands, of pixels. Other fields of science, such as fluorescence microscopy studies of single molecules, should also benefit from this new detector technology.

The ultimate photon detector, however, is not yet at hand. Some challenges remain for Day and colleagues' design¹: for example, although there is very little noise in the readout system, the detector itself has not yet achieved that low level of noise. Whether the noise source is intrinsic or extrinsic is not clear, but I am optimistic that it can be remedied. Meanwhile, in the race to produce the ultimate photon detector, the other candidates (mentioned above) are advancing^{11–13}. The winner — or winners — of this race will be determined in part by issues that go beyond performance, such as cost and ease of use. Practical engineering factors hold sway in astronomy, just as in regular life. ■

Daniel E. Prober is in the Departments of Applied

Physics and Physics, Yale University, New Haven, Connecticut 06520-8284, USA.
e-mail: daniel.prober@yale.edu

1. Day, P. K., LeDuc, H. G., Mazin, B. A., Vayonakis, A. & Zmuidzinas, J. *Nature* **425**, 817–821 (2003).
2. Moseley, S. H., Mather, J. C. & McCammon, D. J. *Appl. Phys.* **56**, 1257–1262 (1984).
3. Stahle, C. K., McCammon, D. & Irwin, K. D. *Phys. Today* **52**, 32–37 (1999).
4. Mather, J. C. *Nature* **401**, 654–655 (1999).

5. Twerenbold, D. *Rep. Prog. Phys.* **59**, 349–426 (1996).
6. Cabrera, B. *et al. Appl. Phys. Lett.* **73**, 735–737 (1998).
7. Peacock, A. *et al. Nature* **381**, 135–137 (1996).
8. Nadis, S. *Science* **274**, 36–38 (1996).
9. Verhoeve, P. *AIP Proc. Conf.* **605** (1), 559–564 (2002).
10. Van Duzer, T. & Turner, C. W. *Principles of Superconductive Devices and Circuits* (Prentice Hall, Englewood Cliffs, New Jersey, 1999).
11. Cabrera, B. *et al. AIP Proc. Conf.* **605** (1), 565–570 (2002).
12. Stevenson, T. R., Pellerano, F. A., Stahle, C. M., Aidala, K. & Schoelkopf, R. J. *Appl. Phys. Lett.* **80**, 3012–3014 (2002).
13. Schoelkopf, R. J. *et al. Science* **280**, 1238–1242 (1998).

Stem cells

Interactive niches

Ihor R. Lemischka and Kateri A. Moore

The microenvironment, or niche, in which stem cells reside controls their renewal and maturation. The niche that regulates blood-forming stem cells in adult animals has eluded researchers — until now.

Cell-fate decisions in the developing embryo are governed by a complex interplay between cell-autonomous signals and stimuli from the surrounding tissue — the microenvironment. Similar processes control the birth and maturation of stem cells that replenish mature cell populations in adults^{1,2}. But where are the stem-cell microenvironments located in adult tissues? And what other cell types contribute to these 'niches'? In mammals, the niches for gut and certain skin stem cells have been pinpointed, and in several cases the molecular signals that emanate from them have been identified^{3–6}. Somewhat ironically, however, the niches that interact with the best-characterized mammalian stem cells, the haematopoietic stem cells (HSCs), which replenish at least ten blood-cell lineages, have been much more elusive. In this issue, Zhang *et al.*⁷ and Calvi *et al.*⁸ now provide insights into the nature of HSC niches in adult animals. These authors demonstrate that osteoblasts — cells that reside in the bone marrow and secrete the calcified bone matrix — have a crucial role in HSC regulation.

Several studies have suggested that primitive HSCs reside next to the inner surface of bone and that they migrate towards the blood vessels at the centre of the bone marrow cavity as they mature and 'differentiate'^{9,10}. Since the 1970s, efforts to characterize the HSC niche have involved developing systems *in vitro* that might mimic some of the features of stem cell–niche interactions *in vivo*¹¹. Indeed, single clones of 'stromal' cells can support HSC self-renewal and differentiation in culture¹². And some of these stromal cell clones are part of the bone-forming 'osteoblastic' lineage, which is consistent with the idea that osteoblasts might be a component of the HSC niche *in vivo*¹³.

Zhang *et al.*⁷ and Calvi *et al.*⁸ have now confirmed that osteoblasts have a function in stem-cell regulation in animals. Both groups employed genetic strategies to increase the

size of the osteoblast population in specific regions of bone. They then looked at how this affected the HSC population. In essence, they found that increasing the number of osteoblasts causes parallel increases in the HSC population.

Zhang *et al.*⁷ looked at the involvement of signalling by bone morphogenetic protein (BMP) in HSC regulation. The activity of BMP is crucial to the development of blood-forming tissue in embryos. The authors show that mutant mice depleted of a cellular receptor for BMP develop bone abnormalities — calcified 'trabecular bone-like areas' form within long bones. And the numbers of primitive long-term (LT) HSCs are approximately doubled in these animals. The authors go to considerable lengths to demonstrate that the HSC pool is specifically increased without concomitant increases in other primitive progenitor cell populations. Such a specific increase in only the LT-HSC population is consistent with very local effects; in other words, it suggests that a very specific niche is functionally enhanced.

To gain further insight, the authors employed a more elaborate genetic strategy in which a fluorescent green protein marker was activated in cells that were depleted of the BMP receptor. Only the osteoblasts that lined the surface of the bone-like area fluoresced. Using a combination of cell markers to label the LT-HSCs, Zhang *et al.* found that the stem cells co-localized with spindle-shaped osteoblasts lining the bone surface. And a doubling of this osteoblast population mirrored the increase in the LT-HSC population in the mutant mice. These osteoblasts, and a subpopulation of the HSCs, expressed N-cadherin, a cell-surface molecule that helps cells adhere to one another. The authors suggest that N-cadherin and a protein that forms a complex with it, β -catenin, might form important components of the interaction between the stem cell and its niche. To prove this, it will be necessary to

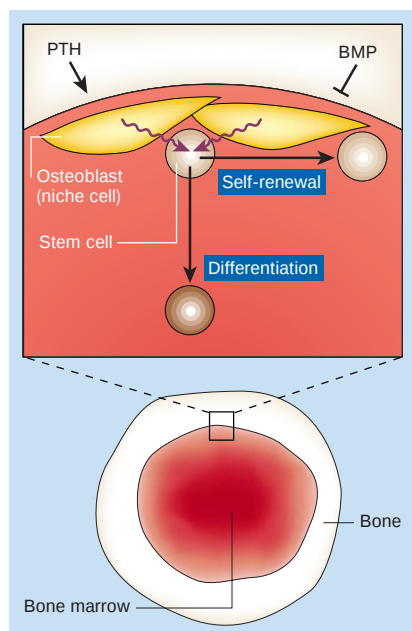


Figure 1 Birth control for stem cells. The niche that regulates the birth and differentiation of blood-forming (haematopoietic) stem cells is formed of osteoblasts (a type of bone-marrow cell) that line the inner surface of bone. Zhang *et al.*⁷ showed that depleting osteoblasts of a receptor for bone morphogenetic protein (BMP) caused a doubling in both the osteoblast population and the stem-cell population. Calvi *et al.*⁸ found a parallel expansion of the stem-cell population when they increased the numbers of osteoblasts by using parathyroid hormone (PTH).

show that N-cadherin is expressed by functional LT-HSCs.

Calvi *et al.*⁸ arrive at some of the same conclusions but with a different approach. These authors also used a genetic strategy to increase osteoblast numbers in genetically engineered mice — but in this case, the osteoblasts of these mutant mice over-produced a cellular receptor that enhanced their growth. The resulting increases in the number of osteoblasts matched increases in the HSC population. Again, the authors showed that only the LT-HSC population increased in size and that other, more committed, progenitor populations were unchanged. To prove that microenvironmental signals caused the HSC population to expand, the authors used an *in vitro* test. They found that cultured stromal cells taken from the genetically engineered animals were better able to support HSCs than were stromal cells taken from normal animals.

Additional experiments implicated the 'Notch' signalling pathway in the expansion of the HSC population in the transgenic animals. This pathway is widely used in many organisms to regulate cell-fate decisions. Osteoblasts from the transgenic mice produced higher than normal levels of Jagged1 — a ligand for the Notch receptor — whereas the

HSCs showed corresponding increases in the levels of the activated intracellular portion of this receptor. Furthermore, an inhibitor of the Notch pathway abrogated the ability of transgenic stromal cells to support HSCs in culture. Finally, when the authors administered parathyroid hormone to normal mice to increase the osteoblast cell population, the HSC population also increased in size. This strategy might prove to have future clinical value.

How precisely is the HSC niche defined by these two studies? It is not yet delineated as clearly as the niches for mammalian gut or skin stem cells, or the embryonic 'germ-line' cells of the fruitfly *Drosophila*. In these systems the structure of the tissue facilitates the identification of stem cells and their immediate progeny. Determining the location of short-term HSC cells and of HSC progeny will be important, as will identifying the molecular signals that emanate from the niche. A fluorescent 'reporter' system has been used to reveal the reception of 'Wnt' signals by HSCs — similar approaches could be used to investigate other signalling pathways, such as the Notch pathway, in the HSC niche¹⁴.

Zhang *et al.*⁷ and Calvi *et al.*⁸ clearly implicate osteoblasts as components of the HSC niche (Fig. 1), but other cell types might also be involved. In addition, it should be kept in

mind that generic terms such as 'osteoblasts', 'endothelial cells' and 'fibroblasts' (other cell types in bone marrow) do not imply that all cells in these categories are identical¹⁵. In future studies it will be important to determine the exact molecular characteristics of the osteoblasts that contact HSCs and thus contribute to the maintenance and renewal of virtually all the cells of the blood.

Ihor R. Lemischka and Kateri A. Moore are in the Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544, USA. e-mail: ilemischka@molbio.princeton.edu

1. Kiger, A. A., Jones, D. L., Schulz, C., Rogers, M. B. & Fuller, M. T. *Science* **294**, 2542–2545 (2001).
2. Spradling, A., Drummond-Barbosa, D. & Kai, T. *Nature* **414**, 98–104 (2001).
3. Kopan, R. *et al.* *Dev. Biol.* **242**, 44–57 (2002).
4. Marshman, E., Booth, C. & Potten, C. S. *BioEssays* **24**, 91–98 (2002).
5. Nishimura, E. K. *et al.* *Nature* **416**, 854–860 (2002).
6. Alonso, L. & Fuchs, E. *Proc. Natl Acad. Sci. USA* **100**, 11830–11835 (2003).
7. Zhang, J. *et al.* *Nature* **425**, 836–841 (2003).
8. Calvi, L. M. *et al.* *Nature* **425**, 841–846 (2003).
9. Lord, B. I. *Int. J. Cell Cloning* **8**, 317–331 (1990).
10. Nilsson, S. K., Johnston, H. M. & Coverdale, J. A. *Blood* **97**, 2293–2299 (2001).
11. Dexter, T. M., Allen, T. D. & Lajtha, L. G. *J. Cell. Physiol.* **91**, 335–344 (1977).
12. Moore, K. A. *Blood* **89**, 4337–4347 (1997).
13. Taichman, R. S. & Emerson, S. G. *Stem Cells* **16**, 7–15 (1998).
14. Reya, T. *et al.* *Nature* **423**, 409–414 (2003).
15. Chi, J.-T. *et al.* *Proc. Natl Acad. Sci. USA* **100**, 10623–10628 (2003).

Metabolism

It's in the genes

Robert W. Furness

Warm-blooded animals of the same species, living in different climates, have different metabolic rates. In birds, this variation is not only due to physiological adaptation — it is inherent in the animals' genes.

Reptiles tend to live in warmer parts of the world because their low metabolic rate hinders life in cold climates. Birds and mammals, on the other hand, have high metabolic rates, which allow them to exploit colder climates by maintaining their body temperatures far above ambient conditions. But many species of bird and mammal are found in both cold and warm climates and metabolism can vary between the different populations. Are these variations driven by physiological adaptations to the local climate, or are they hard-wired into the animals' genetic make-up? Writing in the *Proceedings of the Royal Society*, Wikelski *et al.*¹ suggest that the metabolic rate of birds is a genetically determined characteristic that can respond to selection pressures inherent in the different regions in which the birds live.

The gross differences in metabolic rate between cold-blooded reptiles and warm-blooded birds and mammals are well known, but the patterns of variation between closely related species, or indeed within individual

species, are less well understood. Wikelski and colleagues¹ have now examined the variation in the metabolic rates of different populations of a single species of bird. The authors made an outstanding choice of study species. They needed a species that could be reared in captivity easily, and that existed as several genetically distinct populations, inhabiting a wide range of climates. They chose to work with stonechats (*Saxicola torquata*; Fig. 1, overleaf), a species of bird with several widely distributed 'races'. They show that stonechats from a tropical climate with little seasonal variation (Kenya) have a consistently low metabolic rate, whereas birds from a cooler but still fairly mild climate (Ireland) have a somewhat higher metabolic rate. And birds from colder climates (Austria and Kazakhstan) have by far the highest metabolic rates.

It was demonstrated several years ago that metabolic rate increases with latitude^{2,3} (a rough proxy for cold climate), but the crucial aspect of Wikelski and colleagues' study is that, from a very early age, all the birds were reared in captivity under the same



Figure 1 A little bird with a lot of variation. Different stonechat populations live in different climates and have different metabolic rates. Wikelski *et al.*¹ show that this variation is inherent in the animal's genes.

conditions. So the variations between birds from the different populations could not be due to physiological adaptations to their respective local climates. Instead, the authors suggest that the differences in metabolic rate are specified by their genes.

What selective pressures shape the genetic make-up of different stonechat populations? Stonechats living in mild climates are year-round residents, whereas in locations with harsh winter climates, the birds are long-distance migrants. So the higher metabolic rate of stonechats from the Austrian and Kazakstan populations could be a consequence of selective pressures associated with migration. Wikelski *et al.* suggest that this is not the case because, energetically, migration is a cheap option compared with thermoregulation. The high metabolic rates of the Austrian and Kazakstan populations are more likely to be selected for by occasional periods of freezing temperatures during the breeding season.

Survival of warm-blooded animals in cold climates is possible only if food consumption is sufficient to meet the energy requirements for maintaining body temperature. But high metabolism has other costs besides increasing appetite — lifespan correlates inversely with heart rate across a wide range of mammal species. Mammals with low metabolic rates (and therefore slow heart rates) tend to live longer. Indeed, the relationship between metabolic rate and senescence is a central issue in the study of longevity⁴. Higher metabolic rates lead to faster cell divisions and a more rapid accumulation of mutations, which might eventually impair cell function and contribute to the ageing process. So the lower metabolism of Kenyan stonechats will allow them to survive better where food is scarce, and might also permit slower senescence. This indeed appears to be a general feature of tropical birds and mammals: the pace of life is slow, but survival is high.

Wikelski and colleagues have looked at how metabolic rate varied between populations, but other researchers have investigated

the variation within single populations. In these studies, between half and two-thirds of the inter-population variation was attributed to individual characteristics that remained unchanged in individuals from year to year⁵. So these studies also hint at a genetic basis for metabolic rate. And seasonal⁶ or between-individual⁷ variations in organ size and other aspects of body composition explained much of the individual variation. It seems that the selective pressures that influence metabolism may be complex, influencing metabolic rate through affecting the density of mitochondria — the energy-producing organelles inside cells — or the size of the muscles or liver.

The genetic basis of metabolic rate has far-reaching implications. For example, it is now apparent that metabolism should be considered when selecting animal populations to reintroduce into particular environments. Recent reintroduction programmes brought Scandinavian red kites to Scotland but Iberian kites to England. Which is the better source population? If Scandinavian kites have a higher metabolic rate than their Iberian relatives, they will be better adapted

to surviving cold conditions, but they will need more food to sustain them.

The new study¹ raises many other questions. Have island populations evolved lower metabolic rates to cope with food scarcity? Have equivalent selection pressures shaped other features, such as insulating plumage or pelt, to match the variations in metabolism? Within a population, do individuals with lower metabolic rates have a slower pace of life? Wikelski *et al.* have demonstrated that metabolic rate is a genetic characteristic that can respond to selective pressure. But, as these questions show, there is a lot more to learn about how the environment shapes the evolution of this characteristic.

Robert W. Furness is at the Institute of Biomedical and Life Sciences, Graham Kerr Building, University of Glasgow, Glasgow G12 8QQ, UK.
e-mail: r.furness@bio.gla.ac.uk

1. Wikelski, M. *et al.* *Proc. R. Soc. Lond. B* doi:10.1098/rspb.2003.2500 (2003).
2. Klaassen, M. *Oecologia* **104**, 424–432 (1995).
3. Bryant, D. M. & Furness, R. W. *Ibis* **137**, 219–226 (1995).
4. Kirkwood, T. B. L. *Nature* **270**, 301–304 (1997).
5. Bech, C., Langseth, I. & Gabrielsen, G. W. *Proc. R. Soc. Lond. B* **266**, 2161–2167 (1999).
6. Bech, C. *et al.* *Comp. Biochem. Physiol. A* **133**, 765–770 (2002).
7. Bech, C. & Ostnes, J. E. *J. Comp. Physiol. B* **169**, 263–270 (1999).

Cancer

A twist in a hedgehog's tale

Matthew P. Scott

The genetics of development can often explain the genesis of cancer. This now seems to be true for cancers of the gut, but the patterns of gene expression in these tumours tell a tale with a twist.

Growth is beautifully orchestrated during normal development to produce animals of certain sizes. Similarly, in adult tissues that constantly regenerate, such as blood, gut and skin, a delicate balance is established between cell death and cell division. But genetic damage can destroy the system of checks and balances, sometimes causing cancer. Indeed, it is becoming increasingly clear that many tumours result from normal developmental processes gone awry. Several types of cancer have been linked to the disruption of the 'hedgehog' signalling pathway, which is crucial to the normal development of many tissues^{1,2}. In this issue, Berman *et al.*³ and Thayer *et al.*⁴ now implicate hedgehog signalling in the genesis of cancers of the digestive system. Unusually, however, the fault does not lie with the molecules that propagate the hedgehog signal — instead, the tumours make too much hedgehog.

The hedgehog protein (Hh) is so called because it was discovered as the trigger for a signalling pathway that organizes the pattern of bristles on fruitfly larvae. But Hh is involved in the development of many organs and tissues in numerous animal species.

After being secreted from cells, Hh binds to a receptor protein on the surface of cells nearby. The receptor then transmits an intracellular signal, often culminating in changes in gene expression. Most of the components that transduce the signal within cells have been conserved during more than half a billion years of evolution and are recognizably similar in insects and mammals.

Two components of the Hh pathway are Patched (Ptc), the cell-surface receptor protein to which the secreted signal can bind, and Smoothened (Smo), an intracellular protein that activates genes in response to the Hh signal. In the absence of Hh, Ptc inhibits the activity of Smo. But when a burst of Hh molecules is secreted by nearby cells, the molecules bind to Ptc, unleashing Smo, which then transmits the signal through a further chain of regulators. The result is the activation of certain genes, some of which encode proteins that trigger growth.

These regulatory relationships are the key to understanding what happens in several types of cancer. If Ptc is not sufficiently active, Smo can be overactive and cells can divide when they should not. The most common cause of the deregulation is genetic

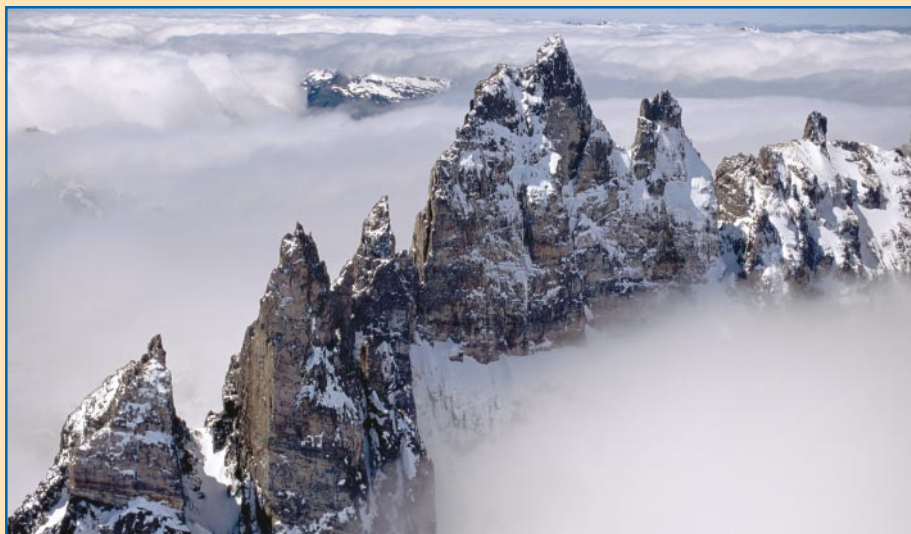
Earth science

How do your mountains grow?

The soaring, jagged peaks of the world's second highest mountain range, the Andes, could owe their stature not to the power of the Earth but to that of the sea. So propose Simon Lamb and Paul Davis in this issue (*Nature* **425**, 792–797; 2003).

Most mountain ranges are created by the collision of two continental plates, such as the grinding action of India against Asia that thrust up the Himalayas. But the Andes are perched on a point where an oceanic plate slips down beneath a continental one. Great mountains aren't usually born of such meetings. Heavy, dense oceanic plates tend to slip underneath continental ones, causing major earthquakes but not world-class mountains. If the push of the mid-Atlantic ridge on tectonic plates were the only factor driving up the Andes, one calculation shows that they would be no more than two kilometres high — half their actual height.

Lamb and Davis argue that the extra push comes from the fact that the forces of the plate collision are focused on a small area — a stretch of the plate boundary where the friction in the trench between the Pacific and South America is particularly high. The cold water current that sweeps up the west



coast of Peru and Chile from high southern latitudes encourages little water evaporation, and therefore little rain. That means there are no rivers to dump worn mountain sediment back into the ocean — sediment that could act as lubrication. Instead, the trench off the coast of the Andes is rough and dry. That extra friction, say the authors, helps to prop up the mountains.

There could even be a positive feedback loop in place, causing the mountains to bulk up more and

more over time. The higher the mountains grow, the drier the coastline becomes — as any wet air coming in from the Atlantic is blocked by the towering peaks — which in turn further reduces erosion and props the mountains up still more.

Complex relationships between rock, air and sea have been found before, although it's usually mountains that are thought to affect climate, rather than the other way around. The Himalayas, for example,

are believed to have changed the air flow enough to spark the formation of the Indian monsoons. And as all that rain weathered away the mountains' rocks, carbon dioxide was taken out of the air and sent down streams as carbonate, to be buried at the bottom of the sea; this extraction of greenhouse gas is thought to have cooled the global climate. But in the Andes, rather than the mountains making the climate, the climate might actually have made the mountains.

Nicola Jones

mutation. Basal-cell carcinoma of the skin — a common result of exposure to sunlight — is most often triggered by ultraviolet damage to both copies of the *ptc* gene^{5,6}. Mutation of *ptc* has also been implicated in certain brain⁷ and muscle tumours⁸. Cancer can also result from mutations in *smo* that convert the protein to a permanently activated form⁹.

Berman *et al.*³ and Thayer *et al.*⁴ now implicate the Hh pathway in the genesis of cancers of the digestive system. Rather little has been understood about the origins and growth of these often fatal forms of cancer, but the role of Hh in the normal development of these tissues hints that defective Hh signalling might allow growth to explode. Curiously, however, it seems that neither *ptc* nor *smo* is mutated in tumours of the digestive tract. Instead, the cancer cells make too much Hh.

Berman *et al.* surveyed the production of Hh in cultured cells derived from several different types of digestive-system tumour, including those of the oesophagus, stomach, biliary tract, pancreas and colon. They detected Hh expression in all of these cell

types, which led them to hypothesize that this signalling molecule might be the trigger for tumour growth. To test this idea they tried to inhibit the growth of cultured tumour cells by using a Hh-blocking antibody. Treatment with the antibody blocked the growth of a wide range of tumour cells, whereas the addition of Hh caused tumour growth to spurt.

Meanwhile, Thayer *et al.* examined Hh production in 20 human pancreatic cancer biopsies and found that the protein was aberrantly expressed in 70% of the specimens. To investigate whether Hh might contribute to the genesis of pancreatic cancer, these authors examined transgenic mice in which Hh had been overproduced in the developing pancreas. All of these mice contained abnormal pancreatic cells that bore similarities to precancerous cells observed in a form of human pancreatic cancer.

Importantly, the link between the Hh pathway and cancer of the digestive tract is accompanied by ideas for treatment. Both Berman *et al.* and Thayer *et al.* inhibited

tumour growth in mice with the drug cyclopamine. This chemical is a teratogen — a compound that can cross the placenta and cause defects in a developing fetus. It was first discovered in extracts from the corn lily, a beautiful but treacherous flower often found in alpine meadows. If a pregnant sheep eats the plant, the fetus develops with cyclopic facial features — the same defect that is caused by inadequate Hh activity in people and mice. It turns out that cyclopamine can block the activity of Smo, so this, or other drugs that affect the Hh pathway, have potential as anti-cancer drugs¹⁰.

Both groups used a 'subcutaneous xenograft model' of digestive-tract cancer to test the effect of cyclopamine. In this model, tumour cells are seeded under the skin of immune-deficient mice. Berman *et al.* waited until the tumours had grown to a certain size, before injecting them with cyclopamine each day. Remarkably, the treated tumours regressed completely within 12 days. Meanwhile, Thayer *et al.* found that treatment with cyclopamine reduced the

growth of subcutaneous 'pancreatic' tumours by about half, but growth was slowed even more if the treatment was begun at the same time that the cells were injected under the skin.

Cancers of the digestive system are often incurable, so the prospect of a new treatment is exciting. But much work remains to be done before we know whether cyclopamine, or any other inhibitors of Hh signalling, such as Hh-blocking antibodies, are effective therapies. Although some of the cell lines tested by Thayer *et al.* were highly susceptible to cyclopamine treatment, others were insensitive to the drug. So some forms of pancreatic cancer might not be linked to defective Hh signalling. Alternatively, some cancers might be caused by mutations in components of the pathway that function downstream of Smo, in which case Smo inhibitors such as cyclopamine would have no effect.

Several other components of the Hh signalling pathway, in addition to Ptc and Smo,

have been identified from studies of fruitfly development. Each molecule in the pathway has the potential to cause cancer if its activity is disrupted or increased. But, as Berman *et al.* and Thayer *et al.* show, revealing the signalling defects that promote tumour growth creates opportunities to devise rational therapies. ■

Matthew P. Scott is in the Departments of Developmental Biology and Genetics, Howard Hughes Medical Institute, Stanford University School of Medicine, California 94305, USA. e-mail: scott@cmgm.stanford.edu

1. McMahon, A. P., Ingham, P. W. & Tablin, C. J. *Curr. Top. Dev. Biol.* **53**, 1–114 (2003).
2. Taipale, J. & Beachy, P. A. *Nature* **411**, 349–354 (2001).
3. Berman, D. M. *et al.* *Nature* **425**, 846–851 (2003).
4. Thayer, S. P. *et al.* *Nature* **425**, 851–856 (2003).
5. Hahn, H. *et al.* *Cell* **85**, 841–851 (1996).
6. Johnson, R. L. *et al.* *Science* **272**, 1668–1671 (1996).
7. Goodrich, L. V. *et al.* *Science* **277**, 1109–1113 (1997).
8. Hahn, H. *et al.* *Nature Med.* **4**, 619–622 (1998).
9. Xie, J. *et al.* *Nature* **391**, 90–92 (1998).
10. Taipale, J. *et al.* *Nature* **406**, 1005–1009 (2000).

Evolution

Ending incongruence

Henry Gee

Recovering the true evolutionary history of any group of organisms has seemed impossible. The availability of large amounts of genomic data promises an era in which the uncertainties are better constrained.

The careful reader of this issue will come across a picture that, at first sight, has a startling message. Those impatient to see it should turn to the report from Rokas *et al.* (A. Rokas, B. L. Williams, N. King & S. B. Carroll *Nature* **425**, 798–804; 2003). The authors' Fig. 4, on page 801, is the object of interest. What, you might ask yourself, is so remarkable about a phylogeny — an evolutionary tree — of seven species of yeast of the genus *Saccharomyces*? Closer inspection, however, will show that the authors are making an unprecedented claim: that this is a fully resolved phylogeny with five internal branches in the tree, each of which has unequivocal support from all the data. For years biologists have tried to find methods to tease evolutionary history from obtuse data. This looks like the best attempt yet.

In principle, it should be possible to describe the phylogeny of a group of organisms as a nested set of bifurcating branches with the appearance of a tree, in which the branching order is a graphic expression of the distribution of features among the organisms concerned. The question that always arises, however, is how one can ever justify the choice of characters that govern branching at each node. Which subset of characters, out of hundreds of possibilities, best reflects the 'true' phylogeny of a group? Even if we can possibly make such decisions,

how should the information from characters be 'weighted'? When technology allowed examination of the sequences of genes, rather than features of anatomy or physiology, the feeling was that truth would be simpler to reach. Genes are a direct expression of inheritance, and not signals refracted through the distorting lens of an organism's physical or biochemical characteristics.

But genes are not immune to external influences: like any feature of anatomy, they have histories that can confuse as well as enlighten. The result has been several decades over which phylogenies of various groups of organism have been based on sequences from one or a few genes, the assumption being that the genes of choice stand reliably for the many thousands of others that remain unsampled. The worry was that no justification could exist for this assumption. But this concern was not often articulated, because little could be done to address the problem.

For many years, single genes were all there were, and people had to make the best of them. The result was endless argument, because a phylogeny supported by analysis of one gene would often be very different from that created using data from another. Retreating into statistical thickets, researchers had to employ more or less sophisticated criteria to decide which of many millions of possible trees was most likely to represent the true history of the group in which they were interested.

The availability of genome-sized quantities of sequence information offers the possibility of changing things. Rokas *et al.* used a database of orthologous genes from seven species of *Saccharomyces*, with a more distant relative, *Candida albicans*, as an 'outgroup'. Orthologues are genes in different species that sequence analysis shows have an ancestral gene in common, while the inclusion of an outgroup is common practice to provide an external standard in phylogenetic analysis. With the database information, Rokas *et al.* could ask just how many genes are needed to produce a reliable phylogeny. By the same token, they could address the issue of why single-gene phylogenies are almost always unreliable.

Rokas *et al.* started with 106 genes represented by orthologues in all their candidate species, and then computed phylogenies using each gene in turn. As had been expected, each gene supported its own particular branching order. In other words, the trees were 'incongruent'. Incongruence can be explained in many ways, but the bottom line — only evident when it is possible to compare large numbers of genes simultaneously — is that there are no identifiable parameters that can predict the performance of genes in any systematic way. But when the researchers combined all the data from these disputatious genes, a *pax genetica* emerged — a single tree that was statistically robust at all points. This is the authors' Fig. 4.

They next considered how much genetic information was needed to recover this phylogeny reliably. The results varied according to the method, and the minimum amount of data required to achieve a single, fully resolved tree will vary according to the nature of the problem being analysed. In the case of the yeasts, however, the 'true' phylogeny could be recovered with remarkably little information. Given that nucleotide sequences in genes do not evolve independently, it was often possible to achieve a better result with small numbers of nucleotides sampled from many genes, rather than whole, single genes.

But hindsight is a wonderful thing, and it would have been interesting to know how confident Rokas *et al.* would have been in their results with minimal data had they not already created a perfectly robust tree with what turned out to be a superabundance of data. This, then, will always be a source of unease. There may be no way to know, without question, how many data are necessary to create a perfectly resolved tree, or indeed if this tree is necessarily the 'true' tree. But evolutionary biologists, like scientists generally, can only ever deal in provisional solutions. What is certain is that the work of Rokas *et al.* has raised the game of phylogenetic reconstruction to a new level. ■

Henry Gee is a senior editor at Nature.

Changing sex at the same relative body size

Similar forces may select for gender switching across taxa in all animals with this facility.

Sex change occurs in a variety of animals, including fish, echinoderms, crustaceans, molluscs and polychaete worms¹. Here we show that the relative timing of sex change is surprisingly invariant across all animals: 91–97% of the variation in size at sex change across species can be explained by the simple rule that individuals change sex when they reach 72% of their maximum size. This suggests that there is a fundamental similarity across all animals, from a 2-mm-long crustacean to a 1.5-m-long fish (Fig. 1), in the underlying forces that select for sex change.

Evolutionary theory suggests that sex change is favoured when the reproductive success (fitness) of an individual varies with its age or size, and the relationship differs between the sexes^{1–4}. In this case, natural selection favours individuals who begin life as the sex whose fitness increases more slowly with age, and then change to the other sex when they are older. Although this idea is well established, the theory is hard to test quantitatively because the underlying trade-offs, such as the relationship between age and fitness, are difficult to measure.

Even without such data, quantitative predictions can be made by using a dimensionless approach⁵. Specifically, the evolutionarily stable age of sex change depends upon several dimensionless properties: k/M , $\alpha \times M$ and δ , where k is relative growth rate (Bertalanffy coefficient), M is adult mortality rate, α is age at first breeding, and δ is a coefficient relating male fertility to size (male fertility correlates with L^δ , where L is size). Populations with the same values of these dimensionless quantities are predicted to have the same relative size at sex change⁵, defined as average size at sex change (L_{50}) divided by maximum size (L_{max}).

Several studies have indicated that k/M and $\alpha \times M$ may be invariant within and even across taxa^{2,6}. Consequently, the same relative size at sex change is predicted whenever δ is also invariant, which might be expected with-

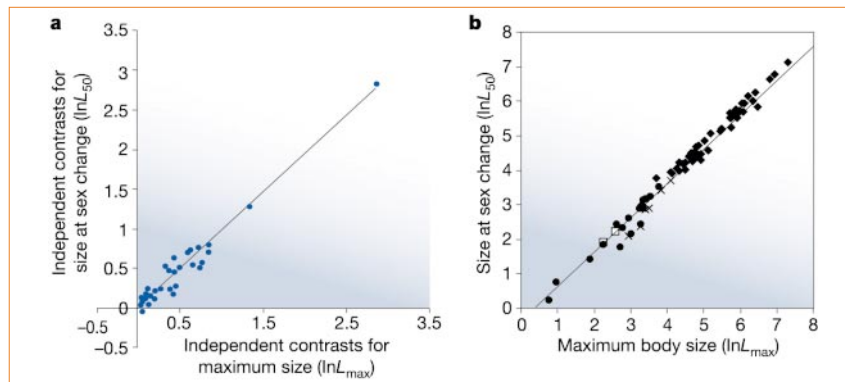


Figure 2 Animals change sex at the same relative body size. **a**, Log-log plot of independent contrasts for L_{50} against L_{max} , where these are the average size and the maximal body size, respectively, at sex change. The slope is fixed at unity and is driven through the origin, as required for the analysis of independent contrasts⁹ ($r^2 = 0.908$, $n = 38$ independent contrasts). The ordinary least-squares (OLS) slope is 0.97 ± 0.05 (95% CI; $r^2 = 0.955$). **b**, Log-log plot of L_{50} against L_{max} for a wide range of sex-changing species, each shown as an independent data point. Data are split by taxa: asterisks, Echinodermata; circles, Crustacea; diamonds, Chordata; crosses, Mollusca. The regression has a slope fixed at unity, giving an intercept of -0.32 ± 0.05 (95% CI; $r^2 = 0.97$, $n = 77$ species). The OLS slope is 1.05 (95% CI) 0.03 , with an intercept of -0.55 (95% CI) 0.07 ($r^2 = 0.98$). The mean relative size at sex change (L_{50}/L_{max}) is 0.72 (95% CI: 0.67 – 0.77), indicating that individuals change sex when they reach 72% of their maximum size. Size (L_{50} and L_{max}) was measured in millimetres before logarithmic transformation.

in species or between closely related species. Consistent with this, an invariant relative size at sex change has been found across populations of a shrimp³ and across fish species⁷.

We investigated the degree of invariance in the relative size at sex change across all sex-changing animals (Fig. 2), using data from 77 species of fish, echinoderms, crustaceans and molluscs. If the relative size at sex change is invariant, then a plot of $\log(L_{50})$ against $\log(L_{max})$ would give a slope of 1.0. We first analysed our data using the method of independent contrasts^{8,9}, based on a composite phylogeny of sex-changing animals. This analysis gave a slope of 0.97 ± 0.05 (95% confidence interval, CI), which is not significantly different from 1.0 ($t_{37} = 1.2$, $P > 0.1$; Fig. 2a). The amount of variance in size at sex change explained by this regression was 95.5%, and remained high (90.8%) when we forced the regression to have a slope of 1.0.

We then analysed our data using species as independent data points. This analysis gave a slope of 1.05 ± 0.03 (95% CI), which is significantly greater than 1.0 ($t_{75} = 3.3$, $P < 0.01$; Fig. 2b). However, this difference reflects the extremely low residual/error variance and is of negligible biological importance, as shown by the fact that forcing the regression to have a slope of 1.0 causes little change in the amount of variation in sex-change size explained by the regression (it dropped by only 1.8%, to 96.7%). Consequently, 91–97% of the variation in mean size at sex change across species can be explained by the simple rule that individuals change sex when they reach 72% (95% CI: 67–77%) of their maximum size.

The amount of variation that can be explained across species in the mean size at sex change is not increased by taking into account life-history variables or taxonomic differences. We tested whether possibly important life-history variables^{1–4,7} and taxonomic groupings could significantly improve the relationship between $\log(L_{50})$ and $\log(L_{max})$.

In possibly important life-history variables^{1–4,7}, there was no significant effect on the direction of sex change (intercept: $F_{1,73} = 1.4$, $P > 0.1$; slope: $F_{1,73} = 0.3$, $P > 0.1$; $n = 77$ species) or for the presence of individuals who mature early as the second sex (termed 'early maturers'; intercept: $F_{1,64} = 0.02$; $P > 0.1$, slope: $F_{1,64} = 0.02$, $P > 0.1$; $n = 68$ species in which the presence or absence of early maturers has been identified). Considering taxonomy, there was no significant difference between the groupings of Chordata, Crustacea, Echinodermata and Mollusca (intercept: $F_{3,69} = 2.7$, $P > 0.05$; slope: $F_{3,69} = 1.9$, $P > 0.1$).

Our results indicate that the relative size of sex change is invariant across all animal species. They suggest that there is a fundamental similarity across animals in the trade-offs and fitness functions associated with sex change. This is remarkable, given the huge variation across species in life-history details such as mating system, sex-change mechanism and size^{1,3,7}. They also suggest a relative invariance in parameters such as $\alpha \times M$ (refs 2, 6), which determine the evolutionarily stable size at sex change⁵.

The problem of when to change sex is formally equivalent to many other problems in life-history theory, such as when individuals



Figure 1 Little and large: the black grouper (1.5 m) and the shrimp *Thor manningi* (2 mm, inset) switch sex at the same relative size.

adjust the sex of their offspring in response to environmental conditions^{1,2,4,10}, suggesting that the dimensionless approach could be applied widely. More generally, our results demonstrate the usefulness of the dimensionless approach, in that it can allow quantitative tests of evolutionary theory in cases where the difficulty of measuring the underlying trade-offs has constrained previous tests to merely qualitative status.

David J. Allsop, Stuart A. West

Institute of Cell, Animal and Population Biology,
University of Edinburgh, King's Buildings,
West Mains Road, Edinburgh EH9 3JT, UK
e-mail: david.allsop@ed.ac.uk

1. Charnov, E. L. *The Theory of Sex Allocation* (Princeton Univ. Press, Princeton, 1982).
2. Charnov, E. L. *Life History Invariants* (Oxford Univ. Press, Oxford, 1993).
3. Warner, R. R. *Trends Ecol. Evol.* **3**, 133–136 (1988).
4. Leigh, E. G. et al. *Proc. Natl Acad. Sci. USA* **73**, 3655–3660 (1976).
5. Charnov, E. L. & Skuladottir, U. *Evol. Ecol. Res.* **2**, 1067–1071 (2000).
6. Gemmill, A. W. et al. *J. Evol. Biol.* **12**, 1148–1156 (1999).
7. Allsop, D. J. & West, S. A. *J. Evol. Biol.* **16**, 921–929 (2003).
8. Felsenstein, J. *Am. Nat.* **125**, 1–15 (1985).
9. Harvey, P. H. & Pagel, M. D. *The Comparative Method in Evolutionary Biology* (Oxford Univ. Press, Oxford, 1991).
10. Frank, S. A. *Foundations of Social Evolution* (Princeton Univ. Press, Princeton, 1998).

Competing financial interests: declared none.

Ancient materials

Analysis of a pharaonic embalming tar

Details of mummification techniques used in dynastic Egypt have emerged from writings in subsequent ancient texts, in which the application of oils (*kedros*, *cedrium*) derived from the cedar tree have been described by Herodotus (490–425 BC)¹ and by Pliny the Elder (AD 23/24–79)². But scholars have since argued that these products were prepared from juniper trees and not from cedar³ — an assertion that is widely accepted by Egyptologists⁴ but which has never been verified by chemical analysis. Here we use gas chromatography to analyse the constituents of a sample of unused entombed embalming material from 1500 BC at a site in Deir el-Bahari, Egypt, and find that its components probably originated from the cedar tree. We also identify one component, guaiacol, as having notable preservative properties.

In ancient Egypt, the deceased were mummified in the hope that this would ensure their eternal survival. The process included removal of the internal organs, followed by desiccation and embalming of the carcass^{5,6}; occasionally cosmetics were applied⁷ as a wealth of different compounds. Active enzymes have recently been isolated from embalmed bones from pharaonic Egypt⁸.

We prepared a methanolic extract from unused embalming material (Fig. 1) found near the mummy 'Saankh-kare', from the eighteenth dynasty, 1500 BC, at Deir el-Bahari. Gas chromatography of the extract showed no

evidence of diterpenoid or triterpenoid resin components, but phenolic compounds (such as cresols, xylenols, guaiacols (2-methoxyphenols)), naphthalenes and azulenes were present (see supplementary information).

The phenolic and naphthalene derivatives probably originated from smouldering wood, with the methoxyphenols arising from lignin-degradation products that resulted from the pyrolysis of soft coniferous wood⁹. Dimethoxyphenols (syringols), on the other hand, are additionally formed by heating the hard wood of deciduous trees¹⁰. The presence of guaiacols without syringols in the embalming material strongly supports an origin from coniferous wood.

The brown embalming resin also contains sesquiterpenoid components, which are normally found in a fluid known as 'cedar oil'. This oil is prepared by extraction with organic solvents of wood from *Cedrus atlantica* and also includes junipene, cadalene, cadinatriene (calamene), cuparene and α -curcumene. Given the prevalence in the resin of guaiacols from coniferous wood-tar oils, our findings indicate that the embalming material was prepared from cedar trees.

The liquid probably originated from the water-containing fraction that is exuded before the tar from dry-distilled cedar wood. In his *Naturalis Historia*, Pliny the Elder² describes the technology: "The wood of the tree is chopped up and put into ovens and heated by means of a fire packed all round outside. The first liquid that exudes flows like water down a pipe; in Syria this is called 'cedar-juice' [in Latin: *cedrium*], and it is so strong that in Egypt it is used for embalming the bodies of the dead." The application of a liquid cedar product is also described by Herodotus¹.

There is an unfortunate tradition of confusion between cedar and juniper trees, which has led to the erroneous assignment of Pliny's *cedrium*. In today's terminology as well as in ancient times, some juniper trees that are not cedar are still called cedar — for example, the American red cedar (*Juniperus virginiana*) and the Mediterranean 'little cedar' (*Juniperus oxycedrus*). In this context, our comparative investigations show that the oils or tars from juniper trees contain high amounts of cedrol or cedrene, respectively, neither of which was found in the *Saankh-kare* embalming material.

The surfaces of mummies embalmed with liquid tars or other liquid resinous materials are essentially free of contaminating micro-organisms. The sealing effect of the embalming agents on the bone surface preserves the enzyme alkaline phosphatase inside for thousands of years⁸. We therefore tested the contribution of phenolic derivatives, which in the embalming agent originate from a smouldering process, and monoterpenes, which occur naturally in resinous balm, to this biochemical conservation process.

Guaiacol, *p*-cymene, limonene and α -pinene were tested for their effect on preserv-

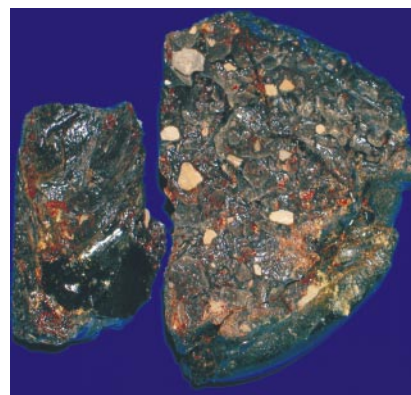


Figure 1 Unused embalming material ('cedrium') entombed with the mummy *Saankh-kare* (from 1500 BC and cemetery field 26225 at Deir el-Bahari in Egypt), at present located at the Egyptian Art Department of the Metropolitan Museum of Art, New York. The dark-coloured parts are organic resin; the pink and lighter-coloured inclusions originate from mineral-based impurities.

ing alkaline phosphatase activity in a model embalming process, in which porcine bones were coated with one of these compounds for 35 days at room temperature. We found that guaiacol was the most effective preservative as the enzyme's specific activity was 12 times higher than that in an untreated bone control and exceeds the activity recovered from bones treated with monoterpenes (cymene has no effect; α -pinene and limonene seem to promote enzyme degradation) or with disinfectants such as phenols (see www.pci.chemie.uni-tuebingen.de/weser/supp_info.htm).

We conclude that liquid tars in general are most efficient in the mummification process. The methyl- and ethylguaiacol identified in the *cedrium* of Deir el-Bahari are characteristic representatives of wood creosotes (from the Greek 'kreas', or flesh, and 'soter', preserver). Creosotes are nowadays used for rapid smoke-drying of meat. Thus, these outstanding conserving properties confirm the statements of Herodotus and Pliny that a "strong" liquid was used for embalming in ancient Egypt.

Johann Koller*, Ursula Baumer*, Yoka Kaup, Mirjam Schmid, Ulrich Weser†

*Bayerische Staatsgemäldesammlungen, Doerner-Institut, 80799 München, Germany

†Physiologisch-Chemisches Institut, Eberhard-Karls Universität Tübingen, 72076 Tübingen, Germany

1. Herodotus *The Histories* (transl. De Selincourt, A.) Book II, 160–161 (Penguin, Harmondsworth, 1954).
2. Pliny (the Elder) *Natural History* (transl. Rackham, H.) Vol. IV, libri XII–XVI (Heinemann, London, 1960).
3. Lucas, A. *J. Egypt. Archaeol.* **XVII**, 13–21 (1931).
4. Griffiths, J. G. A. *The Analyst* **62**, 703–707 (1937).
5. Germer, R. *Das Geheimnis der Mumie* (Artemis, Munich, 1991).
6. Forbes, R. J. *Studies in Ancient Technology* Vol. 3, 2nd edn 190–201 (Brill, Leiden, 1965).
7. David, A. R. in *Ancient Egyptian Materials and Technology* (eds Nicholson, P. T. & Shaw, I.) 372–389 (Cambridge Univ. Press, Cambridge, 2000).
8. Kaup, Y. et al. *J. Am. Res. Center Egypt* **38**, 115–132 (2001).
9. Faix, O. et al. *Holz als Roh- und Werkstoff* **48**, 281–285 (1990).
10. Toth, L. & Wittkowski, R. *Chemie und Zeit* **19**, 48–58 (1985).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.

The nature of human altruism

Ernst Fehr & Urs Fischbacher

University of Zürich, Institute for Empirical Research in Economics, Blümlisalpstrasse 10, CH-8006 Zürich, Switzerland

Some of the most fundamental questions concerning our evolutionary origins, our social relations, and the organization of society are centred around issues of altruism and selfishness. Experimental evidence indicates that human altruism is a powerful force and is unique in the animal world. However, there is much individual heterogeneity and the interaction between altruists and selfish individuals is vital to human cooperation. Depending on the environment, a minority of altruists can force a majority of selfish individuals to cooperate or, conversely, a few egoists can induce a large number of altruists to defect. Current gene-based evolutionary theories cannot explain important patterns of human altruism, pointing towards the importance of both theories of cultural evolution as well as gene–culture co-evolution.

Human societies represent a huge anomaly in the animal world¹. They are based on a detailed division of labour and cooperation between genetically unrelated individuals in large groups. This is obviously true for modern societies with their large organizations and nation states, but it also holds for hunter-gatherers, who typically have dense networks of exchange relations and practise sophisticated forms of food-sharing, cooperative hunting, and collective warfare^{2,3}. In contrast, most animal species exhibit little division of labour and cooperation is limited to small groups. Even in other primate societies, cooperation is orders of magnitude less developed than it is among humans, despite our close, common ancestry. Exceptions are social insects such as ants and bees, or the naked mole rat; however, their cooperation is based on a substantial amount of genetic relatedness.

Why are humans so unusual among animals in this respect? We propose that quantitatively, and probably even qualitatively, unique patterns of human altruism provide the answer to this question. Human altruism goes far beyond that which has been observed in the animal world. Among animals, fitness-reducing acts that confer fitness benefits on other individuals are largely restricted to kin groups; despite several decades of research, evidence for reciprocal altruism in pair-wise repeated encounters^{4,5} remains scarce^{6–8}. Likewise, there is little evidence so far that individual reputation building affects cooperation in animals, which contrasts strongly with what we find in humans. If we randomly pick two human strangers from a modern society and give them the chance to engage in repeated anonymous exchanges in a laboratory experiment, there is a high probability that reciprocally altruistic behaviour will emerge spontaneously^{9,10}.

However, human altruism extends far beyond reciprocal altruism and reputation-based cooperation, taking the form of strong reciprocity^{11,12}. Strong reciprocity is a combination of altruistic rewarding, which is a predisposition to reward others for cooperative, norm-abiding behaviours, and altruistic punishment, which is a propensity to impose sanctions on others for norm violations. Strong reciprocators bear the cost of rewarding or punishing even if they gain no individual economic benefit whatsoever from their acts. In contrast, reciprocal altruists, as they have been defined in the biological literature^{4,5}, reward and punish only if this is in their long-term self-interest. Strong reciprocity thus constitutes a powerful incentive for cooperation even in non-repeated interactions and when reputation gains are absent, because strong reciprocators will reward those who cooperate and punish those who defect.

The first part of this review is devoted to the experimental evidence documenting the relative importance of repeated encounters, reputation formation, and strong reciprocity in human altruism. Throughout the paper we rely on a behavioural—in contrast to

a psychological¹³—definition of altruism as being costly acts that confer economic benefits on other individuals. The role of kinship in human altruism is not discussed because it is well-known that humans share kin-driven altruism with many other animals^{14,15}. We will show that the interaction between selfish and strongly reciprocal individuals is essential for understanding of human cooperation. We identify conditions under which selfish individuals trigger the breakdown of cooperation, and conditions under which strongly reciprocal individuals have the power to ensure widespread cooperation. Next we discuss the limits of human altruism that arise from the costs of altruistic acts. Finally, we discuss the evolutionary origins of the different forms of human altruism. We are particularly interested in whether current evolutionary models can explain why humans, but not other animals, exhibit large-scale cooperation among genetically unrelated individuals, and to what extent the evidence supports the key aspects of these models.

Proximate patterns

Altruistic behaviour in real-life circumstances can almost always be attributed to different motives. Therefore, sound knowledge about the specific motives behind altruistic acts predominantly stems from laboratory experiments. In the following, we first discuss experiments in which interactions among kin, repeated encounters, and reputation formation have been ruled out. Next, we document how the possibility of future encounters and individual reputation formation changes subjects' behaviour. In all experiments discussed below, real money, sometimes up to three months' income^{16–18}, was at stake. Subjects never knew the personal identities of those with whom they interacted and they had full knowledge about the structure of the experiment—the available sequence of actions and the prevailing information conditions. If, for example, the experiment ruled out future encounters between the same individuals, subjects were fully informed about this. To rule out any kind of social pressure, the design of the experiment even ensured in several instances that the experimenter could not observe subjects' individual actions but only the statistical distribution of actions^{19,20}.

Altruistic punishment

The ultimatum game²¹ nicely illustrates that a sizeable number of people from a wide variety of cultures^{22,23} even when facing high monetary stakes^{16,17}, are willing to punish others at a cost to themselves to prevent unfair outcomes or to sanction unfair behaviour. In this game, two subjects have to agree on the division of a fixed sum of money. Person A, the proposer, can make exactly one proposal of how to divide the money. Then person B, the responder, can accept or reject the proposed division. In the case of rejection, both receive nothing, whereas in the case of acceptance, the proposal is implemented. A robust result in this experiment is

that proposals giving the responder shares below 25% of the available money are rejected with a very high probability. This shows that responders do not behave to maximize self-interest, because a selfish responder would accept any positive share. In general, the motive indicated for the rejection of positive, yet 'low', offers is that responders view them as unfair. Most proposers seem to understand that low offers will be rejected. Therefore, the equal split is often the modal offer in the ultimatum game. The decisive role of rejections is indicated by the dictator game, in which the proposer unilaterally dictates the division of the money because the responder cannot reject the offer. The average amount given to the responders in the dictator game is much lower than that in the ultimatum game^{20,24}.

Rejections in the ultimatum game can be viewed as altruistic acts because most people view the equal split as the fair outcome. Thus, a rejection of a low offer is costly for the responder and it punishes the proposer for the violation of a social norm. As a consequence, the proposer is likely to obey the norm in the future by making less greedy offers. For the purpose of this review, we ran an experiment with ten proposers who met a different responder in ten successive rounds. We observed that proposers who experienced a rejection in the previous round increased their offers in the current round by 7% of the available sum of money.

In the ultimatum game, the proposer's action directly affects the responder. However, a key element of the enforcement of many social norms, such as food-sharing norms in hunter-gatherer societies^{2,3}, is that people punish norm violators not for what they did to the punisher but for what they did to others^{25,26}. Norm enforcement involves the punishment of norm violations even by those who are not economically affected by the violation. To study this question experimentally, we conducted a third-party punishment game involving three subjects—an allocator, a recipient, and a third party.²⁷ The allocator is endowed with 100 monetary units (MUs), the recipient has no endowment, and the third party is endowed with 50 MUs. The allocator is free to give whatever he wants to the 'poor' recipient. After the third party has been informed about the allocator's transfer to the recipient, he can spend money to punish the allocator. Every MU spent on punishment reduces the allocator's income by three MUs. Because it is costly to punish, no selfish third party will ever punish. But if a fairness norm applies to the situation, altruistic punishers are expected to punish unfair transfers. In fact, 55% of the third parties punish the allocator for transfers below 50, and the lower the

transfer, the higher the punishment (Fig. 1). Moreover, between 70 and 80% of the recipients expect that allocators will be punished for unfairly low transfers. Similar results have been observed when third parties are given the chance to punish subjects in a 'prisoners' dilemma'²⁷. In this case, they frequently punish a defector if his opponent cooperated. If it is anticipated, the punishment by third parties thus deters non-cooperation.

Altruistic rewarding

Sequentially played social dilemmas are a powerful tool for the study of altruistic rewarding. They come in various forms—as gift exchange games²⁸, trust games²⁹ or sequentially played prisoners' dilemmas³⁰—but the basic structure is captured by the following example. There is a truster and a trustee, both of whom are endowed with 10 MUs. First, the truster decides how many, if any, MUs to transfer to the trustee. Then the trustee decides how much of his endowment to send to the truster. The experimenter doubles any amount sent to the other subject so that, collectively, the two subjects are best off if both transfer their whole endowment: if both keep what they have, each one earns 10; if both transfer their whole endowment, each earns 20. However, a selfish trustee will transfer nothing regardless of how much he received and, therefore, a selfish truster who anticipates this behaviour will never transfer anything in the first place.

This experiment mimics the essence of a vast number of real-life situations. A similar structure characterizes any sequential exchange that takes place in the absence of contracts that are enforced by the courts. In these situations, both players are better off exchanging their goods and favours but there is also a strong temptation to cheat. Despite the incentive to cheat, however, more than 50% of the trustees transfer money and their transfers are the higher the more the truster transferred initially^{28–30}. Like altruistic punishment, the presence of altruistic rewarding has also been documented in many different countries³¹, in populations with varying demographic characteristics³², and under stake levels approaching 2–3 months' income¹⁸.

Strong reciprocity and multilateral cooperation

A decisive feature of hunter-gatherer societies is that cooperation is not restricted to bilateral interactions. Food-sharing, cooperative hunting, and warfare involve large groups of dozens or hundreds of individuals¹. To what extent does strong reciprocity contribute to cooperation in public goods situations involving larger groups of individuals? By definition, a public good can be consumed by every group member regardless of the member's contribution to the good. Therefore, each member has an incentive to free-ride on the contributions of others. Altruistic rewarding in this situation implies that an individual's contributions increase if the expected contributions from the other group members increase. Individuals reward others if the latter are expected to raise their cooperation.

In public goods experiments that are played only once, subjects typically contribute between 40 and 60% of their endowment, although selfish individuals are predicted to contribute nothing³³. There is also strong evidence that higher expectations about others' contributions induce individual subjects to contribute more^{33–35}. Cooperation is, however, rarely stable and deteriorates to rather low levels if the game is played repeatedly (and anonymously) for ten rounds^{36,37}.

The most plausible interpretation of the decay of cooperation is based on the fact that a large percentage of the subjects are strong reciprocators but that there are also many total free-riders who never contribute anything³⁵. Owing to the existence of strong reciprocators, the 'average' subject increases his contribution levels in response to expected increases in the average contribution of other group members. Yet, owing to the existence of selfish subjects, the intercept and the steepness of this relationship is insufficient to establish an equilibrium with high cooperation (Fig. 2). In round

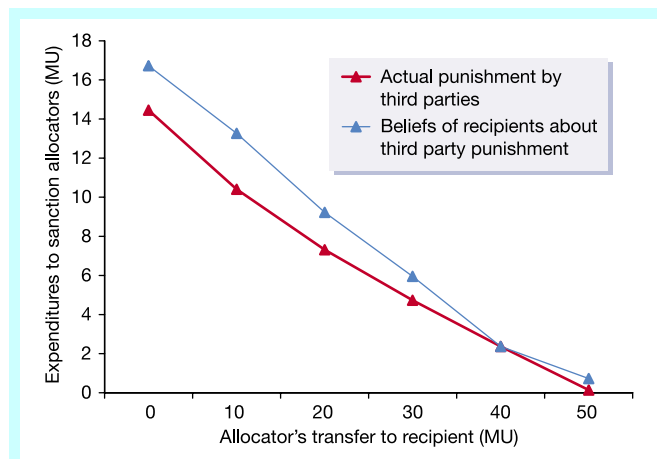


Figure 1 Altruistic punishment by third parties who are not directly affected by the violation of a fairness norm (based on ref. 27). The fair transfer level is given by 50 MUs. The more the allocator's transfer falls short of the fair level of 50 MUs, the more third parties punish the allocator. The recipients of the transfer also expect that the allocators will be punished for unfair transfers.

one, subjects typically have optimistic expectations about others' cooperation but, given the aggregate pattern of behaviours, this expectation will necessarily be disappointed, leading to a breakdown of cooperation over time.

This breakdown of cooperation provides an important lesson. Despite the fact that there are a large number of strong reciprocators, they cannot prevent the decay of cooperation under these circumstances. In fact, it can be shown theoretically that in a population with a clear majority of strong reciprocators, a small minority of selfish individuals suffices to render zero cooperation the unique equilibrium³⁸. This implies that it is not possible to infer the absence of altruistic individuals from a situation in which we observe little cooperation. If strong reciprocators believe that no one else will cooperate, they will also not cooperate. To maintain cooperation in n -person interactions, the upholding of the belief that all or most members of the group will cooperate is thus decisive.

Any mechanism that generates such a belief has to provide cooperation incentives for the selfish individuals. The punishment of non-cooperators in repeated interactions^{39–41} or altruistic punishment^{27,42} provide two such possibilities. If cooperators have the opportunity to target their punishment directly towards those who defect they impose strong sanctions on the defectors. Thus, in the presence of targeted punishment opportunities, strong reciprocators are capable of enforcing widespread cooperation by deterring potential non-cooperators^{39,40,42}. In fact, it can be shown theoretically that even a minority of strong reciprocators suffices to discipline a majority of selfish individuals when direct punishment is possible³⁸.

Repeated interactions and reputation formation

A reputation for behaving altruistically is another powerful mechanism for the enforcement of cooperation in public goods situations. If people are engaged in bilateral encounters as well as in n -person public goods interactions, a defection in the public goods situation, if known by others, may decrease others' willingness to help in bilateral interactions⁴³. Suppose that after each round of interaction

in a public goods experiment, subjects play an indirect reciprocity game⁴⁴. In this game, subjects are matched in pairs and one subject is randomly placed in the role of a donor and the other in the role of a recipient. The donor can help the recipient, and the donor's costs of helping are lower than the benefits for the recipient. The recipient's reputation is established by his decision in the previous public goods round and his history of helping decisions in the indirect reciprocity game. It turns out that the recipients' reputations in the public goods game are an important determinant for the donors' decisions. Donors punish the recipients by being significantly less likely to help when the recipients defected in the previous public goods game. This, in turn, has powerful cooperation-enhancing effects in the future rounds of the public goods game.

Helping behaviour in indirect reciprocity experiments has also been documented in the absence of interactions in public goods games^{45,46}. A crucial element in these experiments is that direct reciprocity is ruled out because no recipient will ever be put in a position where he can give to one of his previous donors. Helping rates between 50 and 90% have been achieved and recipients with a history of generous helping decisions are significantly more likely to receive help themselves. This suggests that the donors' behaviour may be driven by the desire to acquire a good reputation. However, it is also possible that altruistic rewarding drives helping behaviour. A recent study examines this question by allowing only half of the subjects in the experiment to acquire a reputation⁴⁷. This means that one can compare the behaviour of donors who cannot gain a reputation with the behaviour of those who can. The data show that both altruistic rewarding and reputation-seeking are powerful determinants of donors' behaviour. Donors who cannot acquire a reputation help in 37% of the cases whereas those who can gain a reputation help in 74% of the cases. These results indicate that humans are very attentive to possibilities of individual reputation formation in the domain of rewarding behaviours. They exhibit a sizeable baseline level of altruistic rewarding, and when given the opportunity to gain a reputation for being generous, helping rates increase strongly. Humans are similarly attentive to the possibility of repeated interactions with the same individual (reciprocal altruism). The cooperation rate is much higher in social dilemmas if subjects know that there is a possibility of meeting the same partners again in future periods¹⁰.

Little is known about repetition and reputation effects in the domain of punishing behaviours. We conducted a series of ten ultimatum games in two conditions for this purpose—a reputation condition and a baseline condition. In both conditions, 10 MUs have to be divided in every period and every proposer is matched with a new responder in each of the ten games. In the reputation condition, the proposers are informed about the current responder's past rejection behaviour, whereas this knowledge is absent in the baseline condition. This means that the responders in the reputation condition can gain an individual reputation for being tough bargainers by rejecting even high offers. A responder who incurs the short-term cost of a rejection can gain the long-term benefits of a 'good' reputation by inducing future proposers to make him better offers. Because this economic benefit is absent in the baseline condition, subjects who understand the logic of reputation formation will exhibit higher acceptance standards in the reputation condition.

In both conditions, the responders indicated an acceptance threshold in each period, that is, the smallest amount they were willing to accept. The results show that when the subjects were in the baseline condition first, the average acceptance threshold was about 3 MUs, whereas if they entered the reputation condition, their thresholds immediately jumped to more than 4 MUs (Fig. 3a). This jump in the thresholds forced the proposers to increase their offers. Similarly, if the reputation condition took place first, the average thresholds immediately decreased when subjects entered the baseline condition (Fig. 3a). Moreover, this change in the average

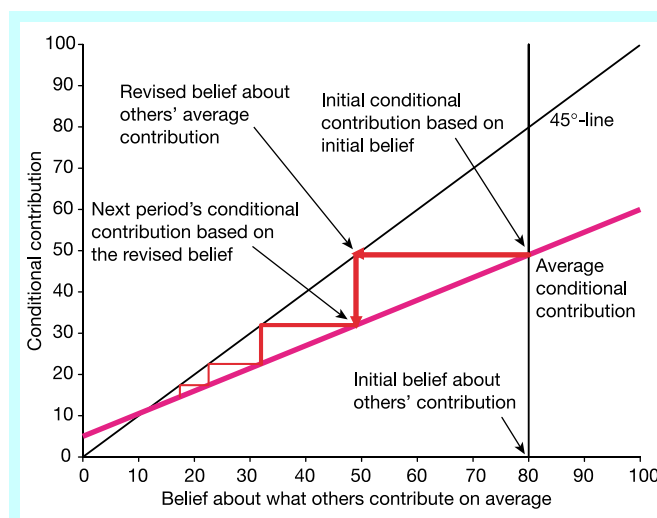


Figure 2 The decay of cooperation over time. Subjects are heterogeneous with regard to their willingness to reward altruistically. This results in the relationship between the expected average contribution of other group members to the public good and the contribution of a representative individual (the average conditional contribution indicated by the purple line). Initially, individuals expect high average contribution rates, say 80% of the endowment. On average, this induces them to contribute 50%. Therefore, expectations are disappointed which leads to a downwards revision of expectations to say, 50% of the endowment. Yet, if individuals expect 50% they will in fact only contribute roughly 30%, causing a further downwards revision of expectations. The process stops at the intersection point with the 45° line, which determines the equilibrium level of altruistic cooperation in this setting.

thresholds is not an artefact of aggregation. It is explained by the fact that the vast majority of responders (82%, $n = 94$) increase the threshold in the reputation condition relative to the baseline (Fig. 3b) while the remaining minority keep the thresholds roughly constant. These results suggest that altruistic punishers clearly understand that if individual reputation building is possible, it pays to acquire a reputation as a tough bargainer. This also means that their rejections in the baseline condition cannot be attributed to cognitive problems in understanding when individual reputation matters and when it does not.

Limits of human altruism

Strongly reciprocal individuals reward and punish in anonymous one-shot interactions. Yet they increase their rewards and punishment in repeated interactions or when their reputation is at stake. This suggests that a combination of altruistic and selfish concerns

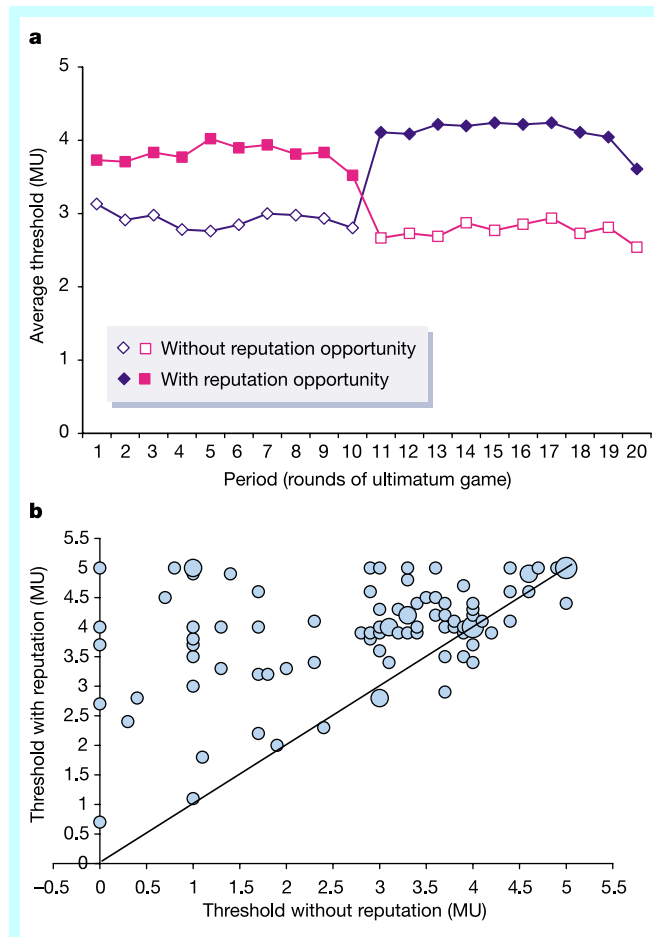


Figure 3 Responders' acceptance thresholds in the ultimatum game with and without reputation opportunities. **a**, Time trend of acceptance thresholds. If the control treatment without the opportunity to build an individual reputation for toughness is conducted first, the responders reject offers below 3 MUs (open blue symbols). Immediately after the implementation of reputation building opportunities in period 11, the acceptance thresholds jump up to more than 4 MUs, indicating the desire to be known as a 'tough' responder (solid blue symbols). If the reputation treatment comes first (purple symbols) the removal of the opportunity to acquire a reputation immediately causes a decrease in responders' acceptance thresholds. **b**, Individual level changes in responders' average acceptance thresholds. The relative size of the circles represents the frequency of observations behind a circle. Responders who increase their average acceptance threshold in the reputation condition relative to the baseline condition generate a data point above the 45° line. The vast majority of responders increase their thresholds when they can gain a reputation for toughness. Only a small minority lowers the thresholds or keeps them roughly constant.

motivates them. Their altruistic motives induce them to cooperate and punish in one-shot interactions and their selfish motives induce them to increase rewards and punishment in repeated interactions or when reputation-building is possible. If this argument is correct, we should also observe that altruistic acts become less frequent as their costs increase. At a higher cost, individuals have to give up more of their own payoff to help others, so that the individuals will exhibit less altruistic behaviour for a given combination of selfish and altruistic motives. The evidence from dictator games and public good games confirms this prediction. If the own payoff that needs to be invested to produce one unit of the public good increases, subjects invest less into the public good^{36,37}. Likewise, if the cost of transferring one MU to the recipient in the dictator game increases, the dictators give less money to the recipients⁴⁸.

Proximate theories

Altruistic rewards and punishment imply that individuals have proximate motives beyond their economic self-interest—their subjective evaluations of economic payoffs differ from the economic payoffs. Although this is an old idea⁴⁹, formal theories of non-selfish motives with predictive power in a wide range of circumstances have only recently been developed. These theories formalize notions of inequity aversion^{38,50} and reciprocal fairness^{51–53}. They predict, for example, that many subjects in the prisoners' dilemma prefer mutual cooperation over unilateral defection, even though it is in their economic self-interest to defect regardless of what the other player does. This prediction is supported by the evidence^{30,54} and has wide-ranging implications. If the players have such preferences, the game is no longer a prisoners' dilemma but an assurance game in which both mutual defection as well as mutual cooperation are equilibria. The crucial point is that such subjects are willing to cooperate if they believe that their opponent will cooperate and, therefore, mutual cooperation is an equilibrium. However, because mutual defection is also an equilibrium, it depends on the individuals' beliefs about the other players' actions as to whether the mutual cooperation or the mutual defection equilibrium is played.

Recent results on the neurobiology of cooperation in the prisoners' dilemma support the view that individuals experience particular subjective rewards from mutual cooperation⁵⁵. If subjects achieve the mutual cooperation outcome with another human subject, the brain's reward circuit (components of the mesolimbic dopamine system including the striatum and the orbitofrontal cortex) is activated relative to a situation in which subjects achieve mutual cooperation with a programmed computer. Moreover, there is also evidence indicating a negative response of the dopamine system if a subject cooperates but the opponent defects.

Evolutionary origins

What are the ultimate origins behind the rich patterns of human altruism described above? It must be emphasized in the context of this question that a convincing explanation of the distinct features of human altruism should be based on capacities which are distinctly human—otherwise there is the risk of merely explaining animal, not human, altruism.

Reciprocal altruism

Reciprocal altruism^{4,5} in the form of tit-for-tat or similar cooperation-sustaining strategies in the repeated prisoners' dilemma is a powerful ultimate explanation for human altruism in small and stable groups. The experimental evidence unambiguously shows that subjects cooperate more in two-person interactions if future interactions are more likely^{9,10}. There are, however, several aspects of human interactions that point towards the need to go beyond reciprocal altruism: first, with a few exceptions^{26,56}, the evolutionary analysis of repeated encounters has been largely restricted to two-person interactions but the human case clearly demands the analysis of larger groups. Unfortunately, the

evolutionary success of tit-for-tat-like strategies of conditional cooperation is extremely limited even in relatively small groups of 4–8 individuals. It turns out that in a repeated n -person prisoners' dilemma, the only conditionally cooperative, evolutionarily stable strategy prescribes cooperation only if all other group members cooperated in the previous period. The basin of attraction of this strategy is extremely small because a few selfish players suffice to undermine the cooperation of the conditional cooperators.⁵⁶

Second, the interacting individuals are forced to stay together for a random number of periods⁶. This assumption is not only violated by many, if not most, animal interactions but it is also clearly violated in the case of humans. Throughout evolutionary history, humans almost always had the option to stop interacting with genetically unrelated individuals. Thus, the choice of cooperation partners has to be modelled explicitly, which brings issues of reputation formation to the forefront of the analysis. Recent experiments indicate that endogenous partner choice and the associated incentives for reputation formation have powerful effects on human cooperation⁵⁷.

Third, reciprocal altruism relies on the idea that altruistic behaviour creates economic benefits for the altruist in the future. Therefore, it has difficulties explaining strongly reciprocal behaviour, which is characterized by the absence of future net benefits. Reciprocal altruism could only explain strong reciprocity if humans behave as if cooperation provides future net benefits, although there are none objectively. The ethnographic evidence suggests that—depending on the interaction partner—humans faced many different probabilities of repeated encounters so that situations often arose in which defection was the best strategy⁵⁸. Indeed, the very fact that humans seem to have excellent cheating detection abilities⁵⁹ suggests that, despite many repeated interactions, cheating has been a major problem throughout human evolution. Therefore, humans' behavioural rules are likely to be fine-tuned to the variations in cheating opportunities, casting doubt on the assumption that humans systematically overestimate the future benefits from current altruistic behaviours.

Reputation-seeking

Evolutionary approaches to reputation-based cooperation^{44,60–64} represent important steps beyond reciprocal altruism. The indirect reciprocity model^{44,60,61} relies on the idea that third parties reward individuals with an altruistic reputation if they can acquire a good reputation themselves by rewarding. It has been shown that aspects of the food-sharing pattern of the Ache of Paraguay can be explained by this logic⁶⁵. The experimental evidence also strongly suggests that a considerable part of human altruism is driven by concerns about reputation. Yet there are still some unsolved theoretical problems that point towards the need for further research. First, the indirect reciprocity approach produces long-run helping rates of roughly 40% if the recipient's benefit is four times the donor's cost, provided that all individuals live in isolated groups without any migration. If, however, genetic mixing between the groups occurs, helping rates decline dramatically and approach zero⁶¹. It would be an important step forward if the indirect reciprocity approach could be modified in such a way that significant helping rates could be maintained under reasonable assumptions about migration between groups. Second, the question of how to model the concept of a good reputation remains open. For example, should an individual who does not help a person with a bad reputation lose his good reputation? Currently the image-scoring approach⁴⁴ gives an affirmative answer to this question while others do not⁶¹.

Third, reputation formation among humans is based on our language capabilities. However, we can use our language to tell the truth or to lie. Thus, what ensures that individuals' reputations provide a reasonably accurate picture of their past behaviours? Fourth, the indirect reciprocity approach is limited to dyadic cooperation. Therefore, it cannot currently explain cooperation in

larger groups. But recent experiments that connect the n -person public good game with an indirect reciprocity game do point towards a potential solution⁴³. Finally, reputation-based approaches cannot account for strong reciprocity unless one assumes that humans behave as if they systematically overestimate the future gains from current altruistic acts—an assumption that is dubious in view of the experimental evidence.

Costly signalling theory also provides a reputation-based ultimate explanation for altruistic behaviour^{62,63}. According to this approach, individuals signal favourable, yet unobservable, traits with altruistic acts, rendering them preferred mating partners or helping in the recruitment of coalition partners in conflicts. The assumption behind this theory is that individuals with better traits have lower marginal signalling costs, that is, lower costs of altruistic acts. Thus, those with better traits are more likely to signal, which allows the inference that those who signal have better traits on average. The advantage of this approach is that it could, in principle, explain contributions to n -person public goods. The weakness is that the signalling of unobservable traits need not occur by altruistic acts but can also take other forms. The approach, therefore, generates multiple equilibria—in some equilibria, signalling occurs via altruistic behaviour; in others, signalling does not involve any altruistic acts. Therefore, this theory has difficulties explaining human altruism unless it is supplemented with some other mechanisms. One such mechanism might be cultural group selection⁶³. If groups where signalling takes place via altruistic behaviour have better survival prospects, selection between groups favours those groups which have settled at a pro-social within-group equilibrium. Since there is no within-group selection against the altruists at the pro-social equilibrium, only weak effects of cultural selection between groups are required here. There is evidence⁶⁶ from Meriam turtle hunters that is consistent with costly signalling theory but so far there is no experimental evidence for altruistic costly signalling.

Gene–culture coevolution

The birth of modern sociobiology is associated with scepticism against genetic group selection⁶⁷; although it is possible in theory, and in spite of a few plausible cases²⁵, genetic group selection has generally been deemed unlikely to occur empirically. The main argument has been that it can at best be relevant in small isolated groups because migration in combination with within-group selection against altruists is a much stronger force than selection between groups. The migration of defectors to groups with a comparatively large number of altruists plus the within-group fitness advantage of defectors quickly removes the genetic differences between groups so that group selection has little effect on the overall selection of altruistic traits⁶⁸. Consistent with this argument, genetic differences between groups in populations of mobile vertebrates such as humans are roughly what one would expect if groups were randomly mixed⁶⁹. Thus, purely genetic group selection is, like the gene-based approaches of reciprocal altruism and indirect reciprocity, unlikely to provide a satisfactory explanation for strong reciprocity and large-scale cooperation among humans. However, the arguments against genetic group selection are far less persuasive when applied to the selection of culturally transmitted traits. Cultural transmission occurs through imitation and teaching, that is, through social learning. There are apparent large differences in cultural practices of different groups around the world and ethnographic evidence indicates that even neighbouring groups are often characterized by very different cultures and institutions⁷⁰. In addition, a culture-based approach makes use of the human capacity to establish and transmit behavioural norms through social learning—a capacity that is quantitatively, and probably even qualitatively, distinctly human^{1,71}.

Recent theoretical models of cultural group selection^{72,73} or of gene–culture coevolution^{71,74} could provide a solution to the puzzle of strong reciprocity and large-scale human cooperation. They are

based on the idea that norms and institutions—such as food-sharing norms or monogamy—are sustained by punishment and decisively weaken the within-group selection against the altruistic trait. If altruistic punishment is ruled out, cultural group selection is not capable of generating cooperation in large groups (Fig. 4). Yet, when punishment of non-cooperators and non-punishers is possible, punishment evolves and cooperation in much larger groups can be maintained⁷³. This is due to the fact that the altruistic punishment of non-cooperators in combination with the imitation of economically successful behaviours prevents the erosion of group differences with regard to the relative frequency of cooperating members. If there are a sufficient number of altruistic punishers, the cooperators do better than the defectors because the latter are punished. Therefore, cooperative behaviour is more likely to be imitated. Moreover, when cooperation in a group is widespread, altruistic punishers have only a small or no within-group disadvantage relative to pure cooperators who do not punish. At the limit, when everybody cooperates, punishers incur no punishment costs at all and thus have no disadvantage. Thus, small cultural group selection effects suffice to overcome the small cost disadvantage of altruistic punishers that arises from the necessity of punishing mutant defectors.

To what extent is there evidence for the role of culture and group selection in human altruism? There is strong evidence from intergenerational ultimatum and trust games that advice from players who previously participated in the experiment increases altruistic punishment and altruistic rewarding⁷⁵. Recent intergenerational public good games where advice is given indicate that later generations achieve significantly higher cooperation levels even in the absence of punishment opportunities⁷⁶. Ultimatum and dictator games with children of different ages show that older children are more generous and more willing to punish altruistically⁷⁷. Although these changes in children's behaviour could be a result of genetic developmental processes, it seems at least as plausible to assume that they are also a product of socialization by parents and peers. Why,

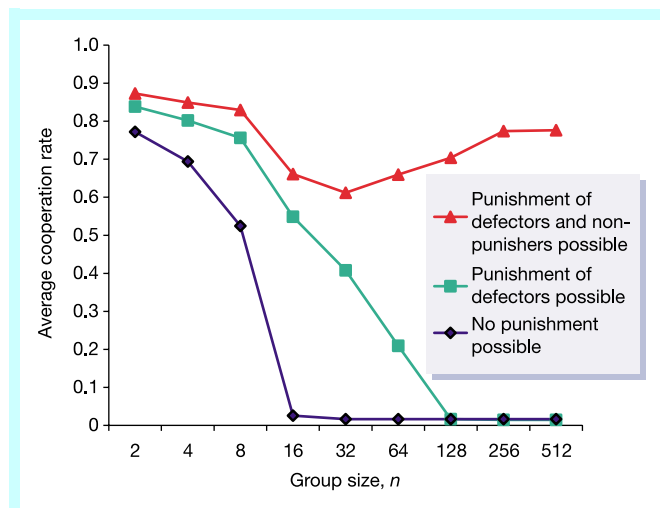


Figure 4 Simulations of the evolution of cooperation in multi-person prisoners' dilemmas with group conflicts and different degrees of altruistic punishment. The simulations are based on the model of ref. 73 but we added the possibility of punishing the non-punishers. There are 64 groups of fixed size n with n ranging from 2 to 512. The figure shows the average cooperation rate in 100 independent simulations over the last 1,000 of 2,000 generations. If the altruistic punishment of defectors is ruled out, cooperation already breaks down for groups of size 16 and larger. If altruistic punishment of defectors is possible, groups of size 32 can still maintain a cooperation rate of 40%. However, the biggest impact from altruistic punishment prevails if non-punishers can also be punished. In this case, even groups of several hundred individuals can establish cooperation rates of between 70 and 80%.

after all, do parents invest so much time and energy into the proper socialization of their children if this effort is futile? Perhaps the strongest evidence for the role of cultural norms comes from a series of experiments in 15 small-scale societies²³, showing decisive differences across societies in the behaviour of proposers and responders in the ultimatum game. Some tribes like the Hazda from Tanzania exhibit a considerable amount of altruistic punishment whereas the Machiguenga from Peru show little concern about fair sharing. Thus, taken together, there is fairly convincing evidence that cultural forces exert a significant impact on human altruism.

Yet, what is the evidence for cultural group selection? There is quite strong evidence that group conflict and warfare were widespread in foraging societies^{78,79}. There are also examples^{70,80} suggesting that group conflict contributes to the cultural extinction of groups because the winning groups force their cultural norms and institutions on the losing groups. However, although these examples are suggestive, they are not conclusive, so further evidence is needed.

If cultural group selection was a significant force in evolution, then the human propensity to reward and punish altruistically should be systematically affected by inter-group conflicts. In particular, altruistic cooperation should be more prevalent if cooperative acts contribute to success in a group conflict. Likewise, people should be more willing to punish defectors if defection occurs in the context of a group conflict. There is evidence from inter-group conflict games indicating that altruistic cooperation in prisoners' dilemmas indeed increases if the game is embedded in an inter-group conflict⁸¹. However, there is no evidence so far showing that inter-group conflicts increase altruistic punishment.

Open questions

We now know a lot more about human altruism than we did one decade ago. There is experimental evidence indicating that repeated interactions, reputation-formation, and strong reciprocity are powerful determinants of human behaviour. There are formal models that capture the subtleties of interactions between selfish and strongly reciprocal individuals, and there is a much better understanding about the nature of the evolutionary forces that probably shaped human altruism. However, there are still a considerable number of open questions. In view of the relevance of cultural evolution, it is necessary to study the relationship between cultural and economic institutions and the prevailing patterns of human altruism. Although recent evidence²³ suggests that market integration and the potential gains from cooperation are important factors, our knowledge is still extremely limited. This limitation is partly due to the fact that far too many experiments use students from developed countries as participants. Instead, we need experiments with participants that are representative of whole countries or cultures and we need to combine behavioural measures of altruism with individual-level demographic data and group-level data about cultural and economic institutions. In view of the theoretical importance of group conflicts and group reputation, much more evidence on how these affect altruistic rewarding and punishment is necessary. We also need more empirical knowledge about the characteristics of the individual reputation acquired by people and how others respond to this reputation.

At the ultimate level, the evolution and role of altruistic rewarding for cooperation in larger groups remains in the dark. Likewise, the empirical study of altruistic rewarding has been largely limited to dyadic interactions and little is known about how cooperation in n -person public good situations is affected if subjects have the opportunity to altruistically reward others after having observed each others' contribution choices. Evolutionary explanations of this kind of altruistic rewarding are likely to be much more difficult than explanations of altruistic punishment because, when cooperation is frequent, rewarding causes high costs for the altruists whereas a credible punishment threat renders actual punishment unnecessary. Finally, to enhance the study of the evolution of human altruism,

there is a great need for empirically testable predictions that are rigorously derived from the evolutionary models. □

doi:10.1038/nature02043.

1. Boyd, R. & Richerson, P. *The Nature of Cultures* (Univ. Chicago Press, Chicago, in the press).
2. Kaplan, H., Hill, J., Lancaster, J. & Hurtado, A. M. A theory of human life history evolution: diet, intelligence, and longevity. *Evol. Anthropol.* **9**, 156–185 (2000).
3. Hill, K. Altruistic cooperation during foraging by the Ache, and the evolved human predisposition to cooperate. *Hum. Nat.* **13**, 105–128 (2002).
4. Trivers, R. L. Evolution of reciprocal altruism. *Q. Rev. Biol.* **46**, 35–57 (1971).
5. Axelrod, R. & Hamilton, W. D. The evolution of cooperation. *Science* **211**, 1390–1396 (1981).
6. Hammerstein, P. in *Genetic and Cultural Evolution of Cooperation. Dahlem Workshop Report 90* (ed. Hammerstein, P.) 1–11 (MIT Press, Cambridge, MA, 2003).
7. Stephens, D. W., McLinn, C. M. & Stevens, J. R. Discounting and reciprocity in an iterated prisoner's dilemma. *Science* **298**, 2216–2218 (2002).
8. Hauser, M. D., Chen, K. M., Frances, C. & Chuang, E. Give unto others: Genetically unrelated cotton-tamarin monkeys preferentially give food to those who altruistically give food back. *Proc. R. Soc. Lond. B* (in the press).
9. Andreoni, J. & Miller, J. Rational cooperation in the finitely repeated prisoner's dilemma: experimental evidence. *Econ. J.* **103**, 570–585 (1993).
10. Gächter, S. & Falk, A. Reputation and reciprocity: consequences for the labour relation. *Scand. J. Econ.* **104**, 1–26 (2002).
11. Fehr, E., Fischbacher, U. & Gächter, S. Strong reciprocity, human cooperation, and the enforcement of social norms. *Hum. Nat.* **13**, 1–25 (2002).
12. Gintis, H. Strong reciprocity and human sociality. *J. Theor. Biol.* **206**, 169–179 (2000).
13. Batson, D. C. *The Altruism Question* (Lawrence Erlbaum Associates, Hillsdale, NJ, 1991).
14. Silk, J. B. Adoption and kinship in Oceania. *Am. Anthropol.* **82**, 799–820 (1980).
15. Daly, M. & Wilson, M. Evolutionary social-psychology and family homicide. *Science* **242**, 519–524 (1988).
16. Cameron, L. A. Raising the stakes in the ultimatum game: Experimental evidence from Indonesia. *Econ. Inq.* **37**, 47–59 (1999).
17. Slonim, R. & Roth, A. E. Learning in high stakes ultimatum games: An experiment in the Slovak republic. *Econometrica* **66**, 569–596 (1998).
18. Fehr, E., Tougareva, E. & Fischbacher, U. *Do High Stakes and Competition Undermine Fairness?* Working Paper 125 (Institute for Empirical Research in Economics, Univ. Zurich, 2002).
19. Bolton, G. & Zwick, R. Anonymity versus punishment in ultimatum bargaining. *Game Econ. Behav.* **10**, 95–121 (1995).
20. Hoffman, E., McCabe, K., Shachat, K. & Smith, V. Preferences, property rights and anonymity in bargaining games. *Game Econ. Behav.* **7**, 346–380 (1994).
21. Güth, W., Schmittberger, R. & Schwartz, B. An experimental analysis of ultimatum bargaining. *J. Econ. Behav. Organ.* **3**, 367–388 (1982).
22. Roth, A., Prasnikar, V., Okuno-Fujiwara, M. & Zamir, S. Bargaining and market behavior in Jerusalem, Ljubljana, Pittsburgh and Tokyo: An experimental study. *Am. Econ. Rev.* **81**, 1068–1095 (1991).
23. Henrich, J. et al. In search of *Homo economicus*: behavioral experiments in 15 small-scale societies. *Am. Econ. Rev.* **91**, 73–78 (2001).
24. Forsythe, R., Horowitz, J. L., Savin, N. E. & Sefton, M. Fairness in simple bargaining experiments. *Game Econ. Behav.* **6**, 347–369 (1994).
25. Sober, E. & Wilson, D. S. *Unto Others—the Evolution and Psychology of Unselfish Behavior* (Harvard Univ. Press, Cambridge, MA, 1998).
26. Bendor, J. & Swistak, P. The evolution of norms. *Am. J. Sociol.* **106**, 1493–1545 (2001).
27. Fehr, E. & Fischbacher, U. Third party punishment and social norms. *Evol. Hum. Behav.* (in the press).
28. Fehr, E., Kirchsteiger, G. & Riedl, A. Does fairness prevent market clearing? An experimental investigation. *Q. J. Econ.* **108**, 437–459 (1993).
29. Berg, J., Dickhaut, J. & McCabe, K. Trust, reciprocity and social history. *Game Econ. Behav.* **10**, 122–142 (1995).
30. Hayashi, N., Ostrom, E., Walker, J. & Yamagishi, T. Reciprocity, trust, and the sense of control—a cross-societal study. *Rational. Soc.* **11**, 27–46 (1999).
31. Buchan, N. R., Croson, R. T. A. & Dawes, R. M. Swift neighbors and persistent strangers: a cross-cultural investigation of trust and reciprocity in social exchange. *Am. J. Sociol.* **108**, 168–206 (2002).
32. Fehr, E., Fischbacher, U., Rosenblatt, B., Schupp, J. & Wagner, G. A nationwide laboratory—examining trust and trustworthiness by integrating behavioural experiments into representative surveys. *Schmoller Jahrbuch* **122**, 519–542 (2002).
33. Dawes, R. M. Social dilemmas. *Annu. Rev. Psychol.* **31**, 169–193 (1980).
34. Messick, D. & Brewer, M. in *Review of Personality and Social Psychology* (ed. Wheeler, L.) (Sage Publ., Beverly Hills, 1983).
35. Fischbacher, U., Gächter, S. & Fehr, E. Are people conditionally cooperative? Evidence from a public goods experiment. *Econ. Lett.* **71**, 397–404 (2001).
36. Isaac, R. M. & Walker, J. M. Group-size effects in public-goods provision—the voluntary contributions mechanism. *Q. J. Econ.* **103**, 179–199 (1988).
37. Ledyard, J. in *Handbook of Experimental Economics* (eds Kagel, J. & Roth, A.) 111–194 (Princeton Univ. Press, 1995).
38. Fehr, E. & Schmidt, K. M. A theory of fairness, competition, and cooperation. *Q. J. Econ.* **114**, 817–868 (1999).
39. Yamagishi, T. The provision of a sanctioning system as a public good. *J. Pers. Soc. Psychol.* **51**, 110–116 (1986).
40. Ostrom, E., Walker, J. & Gardner, R. Covenants with and without a sword: self-governance is possible. *Am. Polit. Sci. Rev.* **86**, 404–417 (1992).
41. Sethi, R. & Somanathan, E. The evolution of social norms in common property resource use. *Am. Econ. Rev.* **86**, 766–788 (1996).
42. Fehr, E. & Gächter, S. Altruistic punishment in humans. *Nature* **415**, 137–140 (2002).
43. Milinski, M., Semmann, D. & Krambeck, H. J. Reputation helps solve the 'tragedy of the commons'. *Nature* **415**, 424–426 (2002).
44. Nowak, M. A. & Sigmund, K. Evolution of indirect reciprocity by image scoring. *Nature* **393**, 573–577 (1998).
45. Wedekind, C. & Milinski, M. Cooperation through image scoring in humans. *Science* **288**, 850–852 (2000).
46. Milinski, M., Semmann, D., Bakker, T. C. M. & Krambeck, H. J. Cooperation through indirect reciprocity: Image scoring or standing strategy? *Proc. R. Soc. Lond. B* **268**, 2495–2501 (2001).
47. Engelmann, D. & Fischbacher, U. *Indirect Reciprocity and Strategic Reputation Building in an Experimental Helping Game* (Working Paper 132, Institute for Empirical Research in Economics, Univ. Zurich, 2002).
48. Andreoni, J. & Miller, J. Giving according to Garp: an experimental test of the consistency of preferences for altruism. *Econometrica* **70**, 737–753 (2002).
49. Thibaut, J. W. & Kelley, H. H. *The Social Psychology of Groups* (Wiley, New York, 1959).
50. Bolton, G. E. & Ockenfels, A. Erc: A theory of equity, reciprocity, and competition. *Am. Econ. Rev.* **90**, 166–193 (2000).
51. Rabin, M. Incorporating fairness into game theory and economics. *Am. Econ. Rev.* **83**, 1281–1302 (1993).
52. Levine, D. K. Modeling altruism and spitefulness in experiments. *Rev. Econ. Dynam.* **1**, 593–622 (1998).
53. Falk, A. & Fischbacher, U. Distributional consequences and intentions in a model of reciprocity. *Ann. Econ. Stat.* **63**, 111–129 (2001).
54. Kiyonari, T., Tanida, S. & Yamagishi, T. Social exchange and reciprocity: confusion or a heuristic. *Evol. Hum. Behav.* **21**, 411–427 (2000).
55. Rilling, J. K. et al. A neural basis for social cooperation. *Neuron* **35**, 395–405 (2002).
56. Boyd, R. & Richerson, P. J. The evolution of reciprocity in sizable groups. *J. Theor. Biol.* **132**, 337–356 (1988).
57. Brown, M., Falk, A. & Fehr, E. Relational contracts and the nature of market interactions. *Econometrica* (in the press).
58. Fehr, E. & Henrich, J. in *Genetic and Cultural Evolution of Cooperation. Dahlem Workshop Report 90* (ed. Hammerstein, P.) 55–82 (MIT Press, Cambridge, MA, 2003).
59. Cosmides, L. & Tooby, J. in *The Adapted Mind* (eds Barkow, J., Cosmides, L. & Tooby, J.) (Oxford Univ. Press, New York, 1992).
60. Alexander, R. D. *The Biology of Moral Systems* (Aldine De Gruyter, New York, 1987).
61. Leimar, O. & Hammerstein, P. Evolution of cooperation through indirect reciprocity. *Proc. R. Soc. Lond. B* **268**, 745–753 (2001).
62. Zahavi, A. Altruism as a handicap—the limitations of kin selection and reciprocity. *J. Avian Biol.* **26**, 1–3 (1995).
63. Gintis, H., Smith, E. A. & Bowles, S. Costly signaling and cooperation. *J. Theor. Biol.* **213**, 103–119 (2001).
64. Nowak, M. A., Page, K. M. & Sigmund, K. Fairness versus reason in the ultimatum game. *Science* **289**, 1773–1775 (2000).
65. Gurven, M., Allen-Arave, W., Hill, K. & Hurtado, M. It's a wonderful life: Signaling generosity among the Ache of Paraguay. *Evol. Hum. Behav.* **21**, 263–282 (2000).
66. Smith, E. A., Blige Bird, R. L. & Bird, D. W. The benefits of costly signalling: Meriam turtle hunters. *Behav. Ecol.* **14**, 116–126 (2003).
67. Williams, G. D. *Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought* (Princeton Univ. Press, Princeton, 1966).
68. Aoki, M. A condition for group selection to prevail over counteracting individual selection. *Evolution* **36**, 832–842 (1982).
69. Long, J. C. The allelic correlation structure of Gainj and Kalam speaking peoples and interpretation of Wright's *f*-statistics. *Genetics* **112**, 629–647 (1986).
70. Kelly, R. C. *The Nuer Conquest: The Structure and Development of an Expansionist System* (Univ. Michigan Press, Ann Arbor, 1985).
71. Bowles, S., Choi, J.-K. & Hopfensitz, A. The co-evolution of individual behaviours and social institutions. *J. Theor. Biol.* (in the press).
72. Henrich, J. & Boyd, R. Why people punish defectors—weak conformist transmission can stabilize costly enforcement of norms in cooperative dilemmas. *J. Theor. Biol.* **208**, 79–89 (2001).
73. Boyd, R., Gintis, H., Bowles, S. & Richerson, P. J. The evolution of altruistic punishment. *Proc. Natl Acad. Sci. USA* **100**, 3531–3535 (2003).
74. Gintis, H. The hitchhiker's guide to altruism: Gene-culture co-evolution and the internalization of norms. *J. Theor. Biol.* **220**, 407–418 (2003).
75. Schotter, A. Decision making with naive advice. *Am. Econ. Rev.* **93**, 196–201 (2003).
76. Chaudhuri, A. & Graziano, S. *Evolution of Conventions in an Experimental Public Goods Game with Private and Public Knowledge of Advice* (Working Paper, Department of Economics, Univ. Auckland, 2003).
77. Harbaugh, W. T., Krause, K. & Liday, S. *Children's Bargaining Behavior: Differences by Age, Gender, and Height* (Working Paper, Department of Economics, Univ. Oregon, 2000).
78. Jorgensen, J. G. *Western Indians: Comparative Environments, Languages, and Cultures of 172 Western American Indian Tribes* (W. H. Freeman, San Francisco, 1980).
79. Otterbein, K. F. *The Evolution of War: A Cross-Cultural Study* (Human Relations Area Files Press, New Haven, 1985).
80. Soltis, J., Boyd, R. & Richerson, P. J. Can group-functional behaviors evolve by cultural-group selection—an empirical test. *Curr. Anthropol.* **36**, 473–494 (1995).
81. Bornstein, G. & Ben-Yossef, M. Cooperation in intergroup and single-group social dilemmas. *J. Exp. Soc. Psychol.* **30**, 52–67 (1994).

Acknowledgements We gratefully acknowledge support by the Ludwig Boltzmann Institute for the Analysis of Economic Growth, by the Swiss National Science Foundation and by the MacArthur Foundation Network on Economic Environments and the Evolution of Individual Preferences and Social Norms. We thank G. Bornstein, S. Bowles, R. Boyd, M. Brewer, J. Carpenter, S. Gächter, H. Gintis, J. Henrich, K. Hill, M. Milinski, P. Richerson, A. Riedl, K. Sigmund, E. A. Smith, D. S. Wilson and T. Yamagishi for comments on the manuscript, and M. Naef, D. Reding and M. Jörg for their research assistance.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to E.F. (efehr@iew.unizh.ch).

Cenozoic climate change as a possible cause for the rise of the Andes

Simon Lamb¹ & Paul Davis²

¹Department of Earth Sciences, Parks Road, Oxford, OX1 3PR, UK

²Department of Earth and Space Sciences, University of California, Los Angeles, California 90095, USA

Causal links between the rise of a large mountain range and climate have often been considered to work in one direction, with significant uplift provoking climate change. Here we propose a mechanism by which Cenozoic climate change could have caused the rise of the Andes. Based on considerations of the force balance in the South American lithosphere, we suggest that the height of, and tectonics in, the Andes are strongly controlled both by shear stresses along the plate interface in the subduction zone and by buoyancy stress contrasts between the trench and highlands, and shear stresses in the subduction zone depend on the amount of subducted sediments. We propose that the dynamics of subduction and mountain-building in this region are controlled by the processes of erosion and sediment deposition, and ultimately climate. In central South America, climate-controlled sediment starvation would then cause high shear stress, focusing the plate boundary stresses that support the high Andes.

The Andes lie in the subducting plate boundary between the Nazca and South American plates, extending for more than 50° of latitude^{1–3} (Fig. 1). They show remarkable systematic variations in both the maximum mean elevation and trench axial depth, with the greatest elevation contrasts between ~10° S and ~33° S (Fig. 2a), reaching ~13 km at ~24° S (ref. 1). In the high Andes, the high heat flow⁴ and widespread late Cenozoic volcanism, together with very low Quaternary strain rates^{5–8} ($<10^{-8} \text{ yr}^{-1}$), virtual absence of large earthquakes⁶, and mixed strike-slip, oblique extension and fold and reverse fault Plio-Quaternary tectonics^{6–13} (Fig. 2) strongly suggest that the lithosphere is weak and buoyancy stress contrasts, caused by elevation and density differences, must nearly be balanced by the available horizontal push in the South American lithosphere. In addition, if crustal shortening is the mechanism of uplift, then this push can never be less than the buoyancy stress in the highlands while they are rising, and it is transmitted across the subduction zone, where it is balanced by the shear stress along the plate interface¹⁴.

Plate interface shear stress along the Andean margin

Figure 2 shows the mean lithospheric plate interface shear stress τ required to balance the buoyancy stress contrasts $\Delta\Gamma$ between the trench and high Andes, plotted as a function of latitude along the Peru–Chile trench. The calculated average shear stresses are in the range 10–50 MPa (for total buoyancy stress contrasts 30–140 MPa), in line with previous estimates of average shear stresses in subduction zones based on heat flow data^{15–17}. At this stage, we do not distinguish between shear stresses in seismic or aseismic portions of the plate interface. The seismically active high-angle normal faults in the vicinity of the Cordillera Blanca at ~10° S (ref. 9)—where both focal mechanisms and the offset of Quaternary moraines point to extension more-or-less orthogonal to the range front^{9,18}, and possibly also extension farther south in the Puna¹⁰—suggest that the calculated shear stresses here may be too high (downward arrows in Fig. 2b). Conversely, reverse fault focal mechanisms in the high Andes at ~15° S, close to where the Nazca ridge is being subducted, suggest that the calculated shear stresses here are too low (upward arrow in Fig. 2b).

If the actual shear stresses everywhere were in fact more-or-less the same, then the high Andes would have to be supporting horizontal deviatoric stresses between 30 and 55 MPa averaged over the entire thickness of the lithosphere, either extensional in the Central Andes, or compressional elsewhere. From the previous

discussion, we do not consider this plausible (Fig. 2a). Also, a pronounced flexural outer bulge between ~15° and 25° S in the Nazca plate¹⁹, adjacent to the trench, suggests that the oceanic lithosphere is bending under higher horizontal compression here than elsewhere along the Andean margin. More evidence comes from a consideration of the thermal structure of the subduction zone.

Thermal modelling shows that the temperature at any depth on the plate interface, T , is proportional to the sum of the heat flux (q_0) from the subducted oceanic plate and shear heating τV (frictional or viscous dissipation) along the plate interface, for subduction at velocity V (refs 15, 16). In this case, given our calculated shear stresses τ (Fig. 2b) and relevant subduction parameters^{3,20–23} (Fig. 2) between 0 and 45° S (Fig. 1), we predict that the thermal structure does not change significantly with latitude, with a nearly constant combined average heat-source term ($q_0 + \tau V$) of ~160 mW m⁻² (Fig. 3) because enhanced shear heating in the high shear stress segments more-or-less compensates for the lower heat flux from older oceanic lithosphere—the blanketing effect of trench sediment may also help to keep this balance (see below). If the down-dip edge of the seismogenic zone is defined by an isotherm²⁴, then the remarkably constant depth of ~50 km to this edge along the Andean margin for the 1995 magnitude $M = 8$ Antofagasta earthquake between 22° and 26° S (ref. 21), as well as pre-1995 seismicity here⁶, and the 1960 $M = 9+$ Concepción earthquake between 36° S and ~45° S (ref. 22), with no clear overall trend elsewhere¹⁷, strongly supports this prediction. The predicted downward heat advection¹⁵ is also consistent with the observed mean forearc heat flow⁴ (~30 mW m⁻²) in northern Chile.

We conclude that all these observations require a focusing of the plate-boundary push along the Andean margin between 10° S and 33° S, because here the buoyancy stress or potential energy per unit volume contrast between the trench and high Andes is higher than elsewhere (Fig. 2b, >100 MPa). This does not depend on other effects, such as enhanced or retarded erosion²⁵, or rheology of the South American lithosphere, because we are considering a vertical and horizontal stress balance in the high Andes where both strain rates and deviatoric stresses are very small, so that rheological effects are negligible.

Trench sediment fill and subduction dynamics

The systematic variations in plate interface shear stresses needed to support the maximum elevation and buoyancy stress contrasts in

the Andes must be a consequence of latitudinal variations in the coefficient of friction or pore fluid pressure on the seismogenic plate interface (normal stress variations are negligible) and/or viscosity or yield strength in the viscous/plastic zone. In view of the discussion above, we believe that the only plausible explanation is some compositional or lithological change in the nature of the plate interface that varies on a length scale of $\sim 1,000$ km—other factors such as the presence of aseismic ridges, the relative plate convergence velocity or direction, or the age of the subducted plate, either do not vary at all significantly, or do not vary in the necessary way or on the appropriate length scale. We believe that the presence or absence of significant trench sediment fill provides the best explanation for lithological or compositional variations in the plate interface on the appropriate length scale.

Along the Andean margin between 10° S and 33.5° S, it is striking that the trench is virtually starved of sediment, with either none or <500 m, where calculated plate interface shear stresses >35 MPa,

but this fill is generally 1–2.5 km thick where shear stresses are much less^{26,27} (Fig. 2b). It is particularly interesting, in this respect, that the zone of seismically active normal faulting in the high Andes at $\sim 10^\circ$ S, where there is a significant decrease in shear stresses, coincides with a marked increase in the thickness of the trench sediment fill²⁶ (Fig. 2b). We recognize, however, that this is merely a snapshot in time, and the trench could be tectonically cleaned out in <1 Myr (ref. 28). We argue (see below) that though trench sediment fill may be episodic, the present sediment distribution is representative of the average fill extending at least into the Pliocene and probably much older²⁸. For example, a longer-term history of subduction erosion in the sediment starved segments, manifested in the extreme trenchward position of early Cenozoic era to Jurassic period volcanic arcs between $\sim 16^\circ$ and 26° S (ref. 29; Fig. 2a), could be attributed to earlier periods of sediment starvation, though it is unclear from the position of the Neogene volcanic arc in the Central Andes whether there has been any net trench migration in the last ~ 25 Myr (ref. 29).

In the sediment-starved sections, where there is no extensive or thick blanketing of the rough oceanic basaltic basement, and tectonic erosion may have been active, significant portions of the plate interface could be between relatively strong basalt in the oceanic plate and relatively dry mafic and felsic basement rocks in the overlying prism¹⁶, with relatively low pore fluid pressures and minimal lubrication from fine-grained subducted trench sediments^{16,30}. All this would promote higher shear stresses compared to segments of the trench where layered turbidite sequences, up to 2.5 km thick, smooth out the top of the subducting plate and provide an abundant source of weak water-rich lubricating material along the plate interface. However, this simple picture has been complicated by the discovery that the sediment-starved margin in northern Chile may be ingesting water-rich sediments³¹. This is because the subduction of rough oceanic topography can erode and undermine the toe of the overriding plate, steepening the slope of the prism and triggering extensive slumps and debris flows which are dragged along the plate interface.

We suggest that the chaotic debris cannibalized from the plate toe does not have the same lubricating effect as the well-stratified and finer-grained trench sediment fill that has smoothed out oceanic basement topography before being subducted. The cannibalized sediment may also be accumulating near the tip of the seismogenic zone, before eventually rising up through the prism as extensive slumping strips off the surface (Fig. 3a)—this may be responsible for limiting the up-dip extension of the seismogenic zone, which is here several kilometres deeper than along the southern Chilean margin³¹. Such a process, as well as reducing, or even preventing, the ingestion of water and trench sediment into deeper parts of the subduction zone (where tectonic erosion of the plate interface may be eating into the basement of the overlying prism), will help cool the up-dip parts of the plate interface by advecting heat upwards, further counteracting the warming and weakening effects of shear heating. Conversely, in sediment-rich trenches, the blanketing of the trench in several kilometres of sediment will raise oceanic basement temperatures near the toe of the overriding prism by up to 100°C (ref. 24, for Andean subduction parameters), which would make up for any thermal deficit from reduced shear heating, without important surface slumping so that sediment is carried much farther down the plate interface (Fig. 3b).

An association between excess trench sediments and great earthquakes³² provides further evidence for the role of trench sediments in subduction dynamics. The two largest subduction earthquakes ever recorded instrumentally—the $M = 9+$ 1960 Concepción and 1964 Alaskan earthquakes—ruptured sediment-rich parts of the trench, with similarly low (10–20 MPa) average plate interface shear stresses³³. The simplest explanation is that the sediment smooths out the stress distribution and lubricates the plate interface, allowing the earthquake to propagate over a wide area of the fault plane

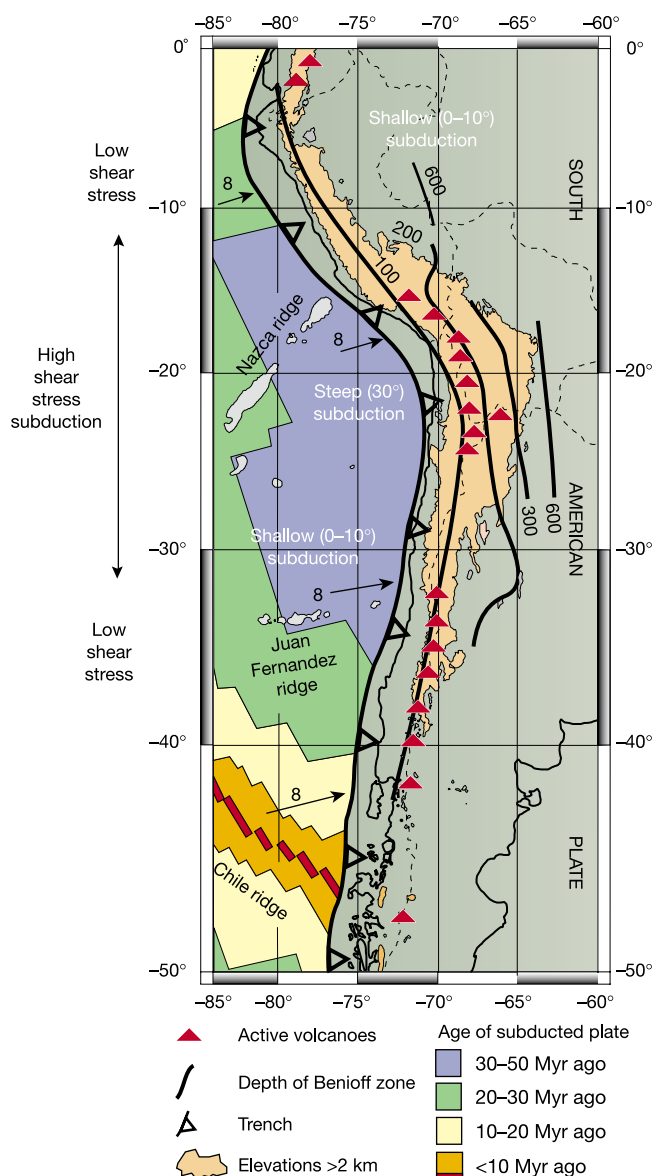


Figure 1 Tectonic map of the Andean plate margin along the western side of South America between 0 and 50° S. Shown are the age and geometry of the subducting Nazca plate^{2,3} and the location of the volcanic arc and regions with elevations $>2,000$ m in the South American plate¹. Depth of Benioff zone is given in km.

once the major asperity barrier is broken³². For sediment-starved trenches the stress distribution is much rougher and earthquakes tend to be smaller³². Clearly, the largest earthquakes do not require the highest average shear stresses—it is the stress coherency that is important. Furthermore the mean stress drop associated with the 1960 Concepción event (1–2 MPa) was small, ~10% of our calculated ambient shear stress²².

The present sediment-starved nature of the Peru–Chile trench is a direct consequence of the arid climate along the west coast of this portion of South America^{26,34,35} (Figs 2a, 4). There is virtually no run-off into the Pacific at these latitudes—the pelagic sedimentary component is likely to be negligible (tens to hundreds of metres) for

this age of oceanic lithosphere and at depths below the carbonate compensation depth. That the arid coastal climate plays such an overriding control is also due to the fact that the Andes form an unbroken watershed for most of their length, so that no drainage from the wetter eastern parts of South America reaches the Pacific coast. However, where the coastal climate is significantly wetter, south of ~31°S, there is some northward transport of sediment along the trench axis by axial turbidite flows, with thicknesses up to 500 m reaching as far north as ~28°S, although the Juan Fernandez ridge may also have acted as a local sediment barrier^{26,27} (Fig. 2a). Likewise, axial sediment flow at low latitudes has transported sediment along the trench as far south as ~12°S (ref. 26) (Fig. 2a).

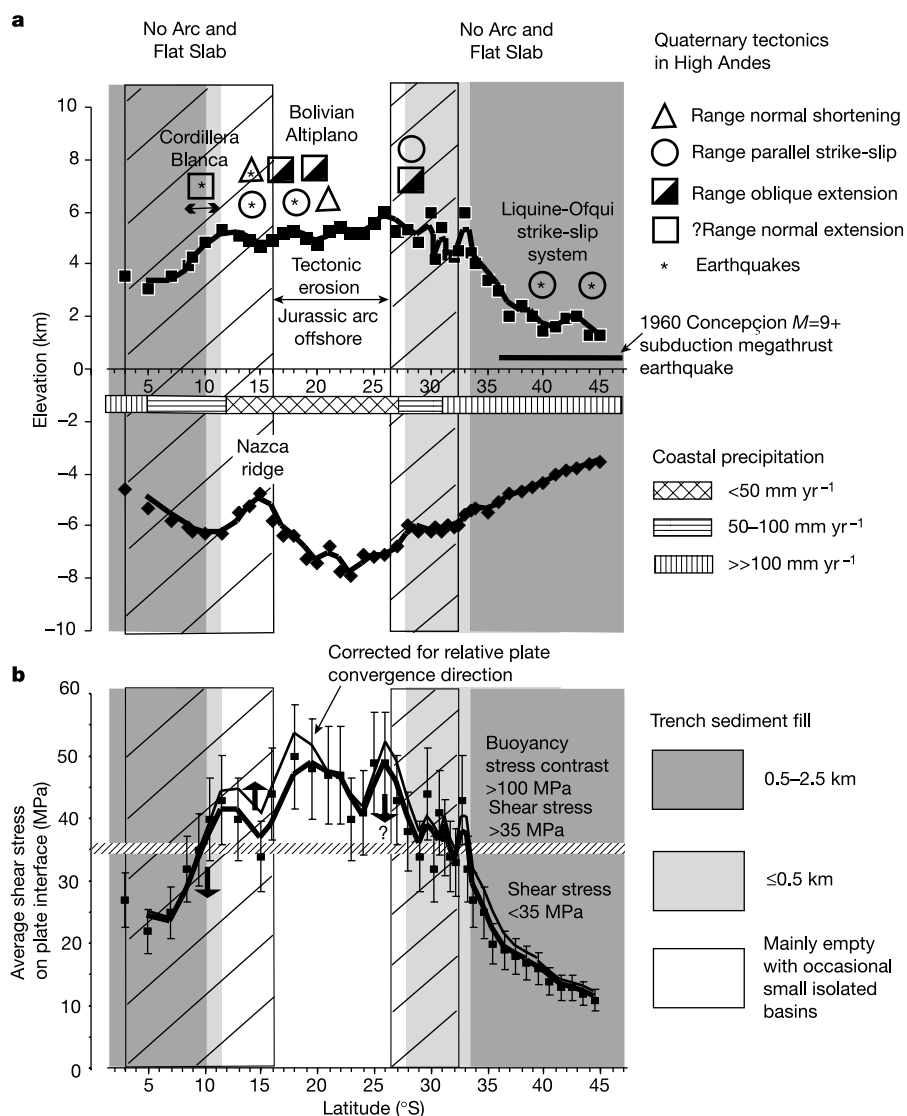


Figure 2 Plots showing latitudinal variations along the length of the Peru–Chile trench between ~3° to 46°S (ref. 1). Information from the Andes has been projected along directions normal to the trench. **a**, Average maximum elevations of high Andes and maximum depth of trench (averaged over 30 km length scale normal to trench, plotted here with a two-point running average curve fit), with the main features of Quaternary tectonics in high Andes, and coastal precipitation. Note the rupture zone of the $M = 9+$ 1960 Concepción earthquake between 36° and 46°S²², evidence for tectonic erosion between 16° and 26°S²⁹, and location of flat slab segments in the subduction zone³. **b**, Shear stresses (τ) on the plate interface¹⁴ required to balance buoyancy stress contrasts ($\Delta\Gamma$) between the trench and high Andes, calculated from a simple one-dimensional analysis across the subduction zone using the method of ref. 51 with

appropriate crustal⁵² and lithospheric parameters ($\rho_{\text{crust}} = 2.85 \text{ g cm}^{-3}$, $\rho_{\text{mantle}} = 3.25 \text{ g cm}^{-3}$, maximum depth of interplate coupling²⁰ = $85 \pm 5 \text{ km}$, reference crustal thickness = $37.5 \pm 5 \text{ km}$, Moho intersection of subduction zone = $45 \pm 5 \text{ km}$, plate interface dip^{3,20–22} = $20 \pm 5^\circ$). This approach is permissible because the width-to-length aspect ratio of the Andean ranges is $\ll 1$ so that along-length gradients of shear stress are small. A small correction for the deviation of subduction from orthogonal²³ ($<30^\circ$) gives an upper acceptable estimate of shear stress. The curve fit is a two-point (100 km scale) running average. Error bars show approximate 1-sigma uncertainties (~17%), each based on 1,000 random trials. Portions of the trench where $\Delta\Gamma > 100 \text{ MPa}$, $\tau > 35 \text{ MPa}$ coincide with either no or very thin (<500 m) axial sediment fill^{26,27}.

Climate change and Andean tectonic evolution

If we are right that the buoyancy stress contrasts in the Andes can only be supported because of the role of climatically controlled trench sediment fill, then it is important to ask whether these elevation contrasts could have existed in the past when the global climate was significantly different. Both the Nazca–South America relative plate convergence history and the inherited rheology of the South American lithosphere have often been invoked as primary controls of the Cenozoic tectonic evolution of the Andes^{5,10,13,14,36–38}, although phases of intense behind-arc shortening seem to correlate better with periods of slow rather than fast convergence^{23,36,37,39} (Fig. 5f)—possibly because slower convergence will result in a cooler plate interface by reducing any shear heating¹⁶.

Inherited rheological effects can certainly influence the rate, amount and distribution of deformation if deviatoric stresses are significant^{5,14}, but unless they lead to a change in the plate interface shear stress or trench-highland buoyancy stress contrast, as previously described, they cannot determine the maximum elevations of the high Andes or the timing of deformation. A wider zone of weaker lithosphere may be part of the explanation for the width and amount of deformation in the Central Andes^{10,14} but it does not explain the higher mountains here, which require higher horizontal normal stresses and more lateral support at the plate interface.

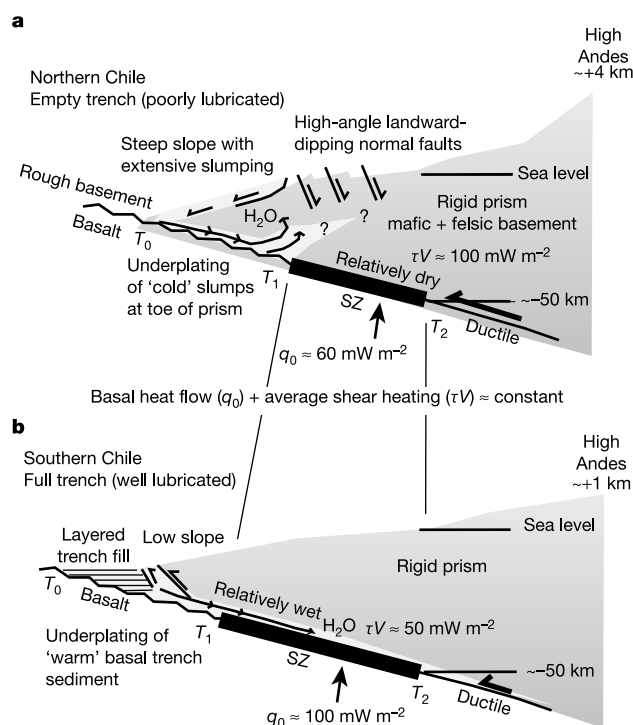


Figure 3 Two cartoons illustrating how the presence or absence of significant trench fill could affect the process of subduction. In both cases, the convergence velocity V , and the combined heat derived both from the subducted plate (q_0) and average shear heating along the plate interface (τV) is more-or-less the same. The seismogenic zone (SZ) is assumed to be defined by isotherms T_1 and T_2 , with T_2 generally lying at ~ 50 km.

a, Extensive undermining of the toe of sediment-starved trenches, such as that in northern Chile, may drag chaotic and coarse debris into the subduction zone, but this probably accumulates and rises near the up-dip edge of the SZ as slumping strips off the overlying prism, helping to cool the plate interface as well as limiting further down-dip movement of the sediment. High-angle landward-dipping normal faults, typical of this type of margin, may be the coastal expression of this. **b**, Trenches full of well-stratified fine-grained sediment may result in 'well-lubricated' and smooth³² subduction zones because the wet sediment may form a weak detachment extending far down the plate interface. Blanketing of the oceanic basement by sediment will also help to warm the subducted oceanic basement.

Indeed, there is no systematic overall correlation between the elevations, amount of deformation and width of the Andes—high elevations are found as far south as 33° S, and as far north as 12° S, where the widths are only a fraction of those around 20° S (compare Figs 1 and 2a). For this reason, given our previous conclusions, we believe that the global cooling of the climate and deep ocean⁴⁰ (Fig. 5a) must be considered as another important effect on the tectonic evolution.

The extreme arid climate of northern Chile and Peru (precipitation $< 50 \text{ mm yr}^{-1}$, see Figs 2a, 4), is a consequence of both atmospheric and ocean circulation, as well as the more local rain shadow effect of the Central Andes themselves, blocking flow of moist air from farther east, in the Amazon region^{35,41}. The cold Peru–Chile current system⁴² (PCC) and wind-driven coastal upwellings³⁵ play an important role in maintaining this aridity by significantly lowering the near-shore air temperature (Fig. 4) and effectively starving any sea-derived winds of moisture, as well as suppressing the rising of and precipitation from the air column³⁵.

An eastern boundary current system, with associated upwellings, has flowed along the western margin of South America since at least the earliest Cenozoic⁴³. However, the temperature of both shallow- and deep-water masses must have been strongly influenced by cooling around Antarctica. This led to enhanced oceanographic and latitudinal temperature gradients⁴⁴ and a northward shift and strengthening of the Antarctic circumpolar current (ACC), resulting in stronger northward advection of cold subpolar water into the southern PCC. In this case, the large expansion of the East (at ~ 14 Myr) and West (at ~ 6 Myr) Antarctic ice sheets, after the mid-Miocene climatic optimum⁴⁰, signal a pronounced cooling of both

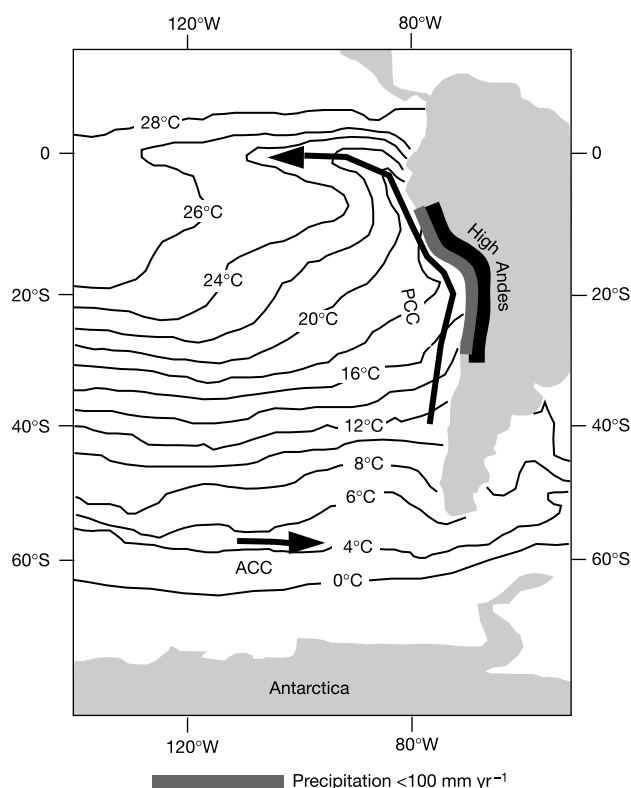


Figure 4 The Peru–Chile current system⁴² and the associated wind-induced oceanic upwelling can be clearly seen in the sea surface temperatures (interpolated satellite and *in situ* measurements from NOAA website: www.nodc.noaa.gov) for July 2002—a tongue of water nearly 8°C colder than that at equivalent latitudes farther west extends up the west coast of South America. This plays an important role in limiting precipitation at mid-latitudes along the coast, where the Andes are highest and widest. The Antarctic circumpolar current flows round Antarctica.

the PCC and deep Pacific water, when the global climate was also cooling (Fig. 5a–c). This seems to correlate remarkably well with the available evidence for the most rapid phase of surface uplift and shortening ($\sim 10 \text{ mm yr}^{-1}$) in the behind-arc region of the central Andes^{5,13,45} (Fig. 5d, e). Intense shortening in the sub-Andean zone initiated in the mid- to late-Miocene (10–20 Myr) with $>100 \text{ km}$ of underthrusting of the Brazilian shield^{13,38,46} (Fig. 5e), and 1–3 km of surface uplift in the Bolivian Altiplano and Eastern Cordillera, so that by the Plio-Pleistocene epoch average elevations exceeded 3 km for the first time^{13,46,47}.

We suggest that the cooling of the PCC (possibly in combination with essentially synchronous planetary cooling^{40,41,44}) provides the causal link between climate change and Andean tectonics, triggering the Mio-Pleistocene trend towards increased aridity along the west coast of South America, which resulted in hyperaridity in the last 3–4 Myr (ref. 41). This aridity was necessary to greatly reduce or cut off the sediment supply to the trench, depriving the plate interface of its lubrication and raising shear stresses to the levels required to push up and support the high Andes. Judging from Fig. 2, these effects seem to become important once the trench fill drops below $\sim 500 \text{ m}$ —this may reflect the sediment required to smooth oceanic basement topography. Strong positive feedbacks between uplift, strengthening of the southeastern Pacific anticyclone, and the

development of a pronounced orographic barrier to moist air from the Amazon basin, would have further promoted the increasing aridity, making this region even more sensitive to the forcing of global climate change.

We speculate that a rise in plate-interface shear stresses could also have occurred in the Eocene, associated with the onset of Cenozoic cooling⁴⁰ (Fig. 5a). The pre-Miocene climate along the west coast of South America is largely unknown. However, the Eocene to Miocene eastward migration of the volcanic arc²⁹ can be explained by sediment starvation and tectonic erosion in the trench. Thus, it is plausible that the late Eocene initiation of intense back-arc shortening in the Bolivian Andes, around $\sim 40 \text{ Myr}$, when the main behind-arc ranges of the central Andes first started to rise^{13,48,49} (Fig. 5e, f), may be a consequence of enhanced shear stresses along the plate interface, caused at least in part by coastal aridity and trench sediment starvation in the Eocene as a consequence of global cooling⁴⁰. Another possible factor is flattening of the slab between ~ 35 and 25 Myr (Fig. 5e), associated with the shutting off of the volcanic arc at this period^{13,48}, but the lack of any systematic relation between the present flat slab regions and Andean topography (Fig. 2a) suggests that this is probably not very significant.

We conclude that high mountain ranges like the central Andes

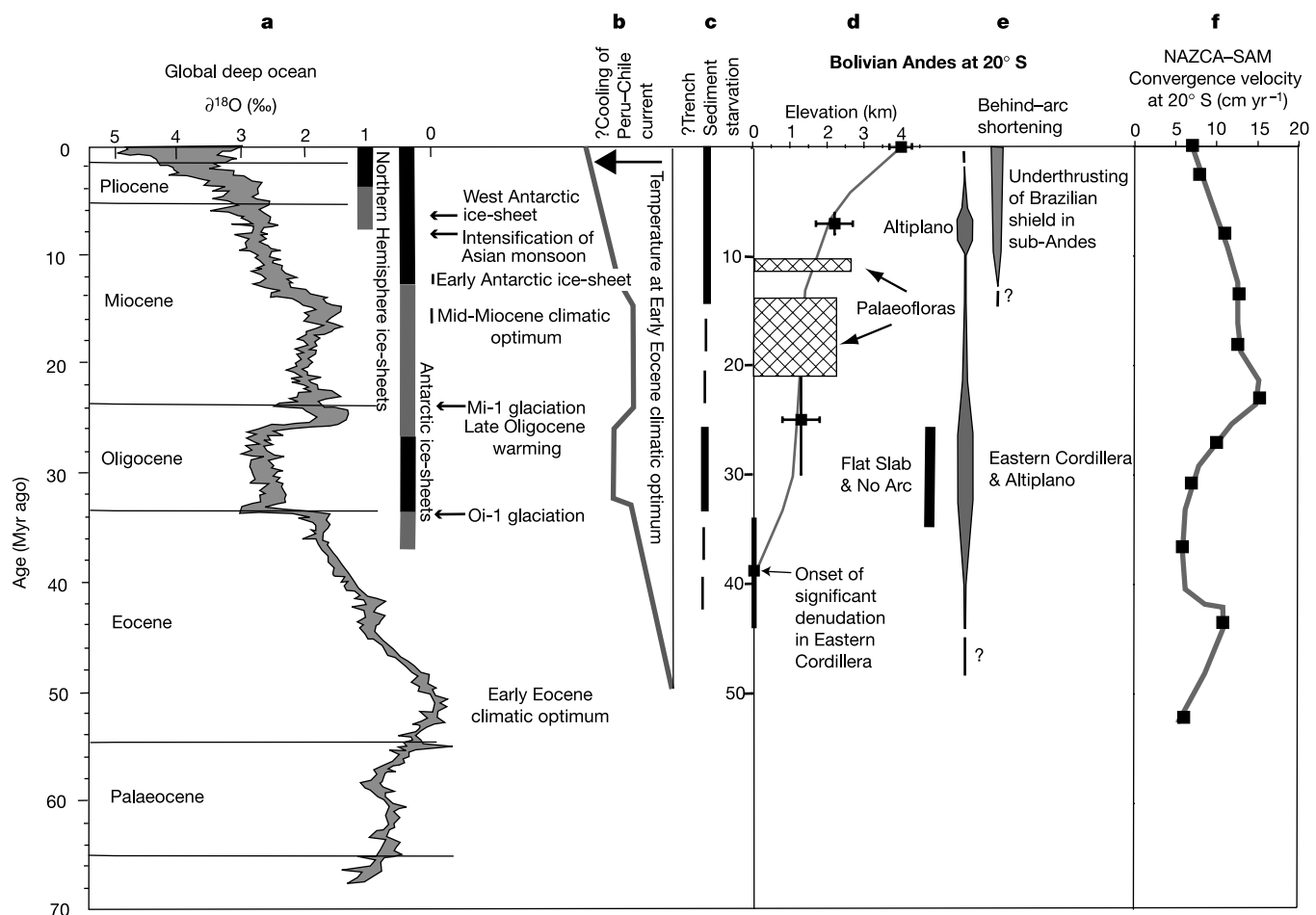


Figure 5 Major global climatic trends, and Andean tectonic evolution at $\sim 20^\circ \text{ S}$, compiled from various sources. **a**, Long-term trends in benthic oxygen isotope ratios⁴⁰, reflecting both cooling of the deep ocean and changes in ice volumes. **b**, Postulated trend in the general surface water temperatures in the Peru–Chile current system. **c**, Postulated phases of sediment starvation in the Peru–Chile trench along the central Andes during periods of high coastal aridity. **d**, **e**, Mean elevation history of the Altiplano and western margin of the Eastern Cordillera in the Bolivian Andes, based on a self-consistent

inversion of estimates from geomorphological data^{13,45} combined with crustal-thickening estimates and their timings^{47–49} from crustal-shortening data^{13,52}, which are broadly in line with palaeobotanical constraints⁴⁷ (shown by cross-ruled boxes), though there are still significant uncertainties. **f**, The absolute relative convergence velocity^{36,37} at 20° S between the South American (SAM) and Nazca plates (NAZCA)—note the slow-down in the last $\sim 25 \text{ Myr}$, and also between 40 and 45 Myr.

may not be typical features of active plate margins, but restricted to regions where climatic conditions are right. The climate here, by promoting or inhibiting sedimentation, may help to focus the available plate-driving forces to relatively restricted portions of subducting plate boundaries, raising the local shear stresses to levels needed to support mountain belts with elevations >3 km. For example, if the Atlantic mid-ocean-ridge push is the main driving force in the Andes, then without some sort of focusing, it would not be expected to support mountains higher than 0.5 to 2.0 km (ref. 50). Finally, an increase in plate interface shear stresses might be a significant driving force for phases of relative plate motion deceleration. □

Received 21 March; accepted 12 September 2003; doi:10.1038/nature02049.

1. Digital Elevation Model (3m × 3m) of South America (Getech, Univ. Leeds, Leeds, UK, 1998).
2. Muller, R. D., Roest, W. R., Royer, J.-Y., Gahagan, L. M. & Sclater, J. G. Digital isochrons of the world's ocean floor. *J. Geophys. Res.* **102**, 3211–3214 (1997).
3. Cahill, T. & Isacks, B. L. Seismicity and shape of the subducted Nazca plate. *J. Geophys. Res.* **97**, 17503–17529 (1992).
4. Springer, M. Heat-flow density across the Central Andean subduction zone. *Tectonophysics* **291**, 123–139 (1998).
5. Lamb, S. H. Active deformation in the Bolivian Andes, South America. *J. Geophys. Res.* **105**, 25627–25653 (2000).
6. National Earthquake Information Center (NEIC) *Earthquake Catalogue* (World Data Centre for Seismology, United States Geological Survey, Denver, 2003).
7. Mercier, J. L. *et al.* Changes in the tectonic regime above a subduction zone of Andean type: the Andes of Peru and Bolivia during the Plio-Pleistocene. *J. Geophys. Res.* **97**, 11945–11982 (1992).
8. Suarez, G., Molnar, P. & Burchfiel, B. C. Seismicity, fault plane solutions, depth of faulting, and active tectonics of the Andes of Peru, Ecuador, and Southern Columbia. *J. Geophys. Res.* **88**, 10403–10428 (1983).
9. Schwartz, D. Paleoseismicity and neotectonics of the Cordillera Blanca Fault Zone, Northern Peruvian Andes. *J. Geophys. Res.* **93**, 4712–4730 (1988).
10. Allmendinger, R. W., Jordan, T. E., Kay, S. M. & Isacks, B. L. The evolution of the Altiplano Plateau of the Central Andes. *Annu. Rev. Earth Planet. Sci.* **25**, 139–174 (1997).
11. Cembrano, J., Herve, F. & Lavenu, A. The Liquine Ofqui fault zone: a long lived intra-arc fault system in southern Chile. *Tectonophysics* **259**, 55–66 (1996).
12. Cladouhos, T., Allmendinger, R., Coira, B. & Ferrar, E. Late Cenozoic deformation in the central Andes: Fault kinematics from the northern Puna, northwestern Argentina and southwestern Bolivia. *J. Soc. Am. Earth Sci.* **7**, 209–228 (1994).
13. Lamb, S. H. & Hoke, L. Origin of the high plateau in the central Andes, Bolivia, South America. *Tectonics* **16**, 623–649 (1997).
14. Lamb, S. H. Vertical axis rotation in the Bolivian orocline, South America, 2: Kinematic and dynamical implications. *J. Geophys. Res.* **106**, 26605–26632 (2001b).
15. Molnar, P. & England, P. Temperatures, heat flux, and frictional stress near major thrust faults. *J. Geophys. Res.* **95**, 4833–4856 (1990).
16. Peacock, S. in *Subduction, Top to Bottom* (eds Bebout, G. E. *et al.*) 119–133 (Geophys. Monogr. 96, American Geophysical Union, Washington, 1996).
17. Tichelaar, B. W. & Ruff, L. J. Depth of seismic coupling along subduction zones. *J. Geophys. Res.* **98**, 2017–2037 (1993).
18. Doser, D. I. The Ancash, Peru, earthquake of 1946 November 10: evidence for low-angle normal faulting in the high Andes of northern Peru. *Geophys. J. R. Astron. Soc.* **91**, 57–71 (1987).
19. Judge, A. V. & McNutt, M. Curvature and elastic plate thickness in the Peru–Chile trench. *J. Geophys. Res.* **96**, 16,625–16,640 (1991).
20. Delouis, B. *et al.* The Andean subduction zone between 22 and 25°S (northern Chile): precise geometry and state of stress. *Tectonophysics* **259**, 67–80 (1996).
21. Klotz, J. *et al.* GPS-derived deformation of the Central Andes including the 1995 Antofagasta Mw = 8.0 Earthquake. *Pure Appl. Geophys.* **154**, 709–730 (1999).
22. Barrientos, S. E. & Ward, S. N. The 1960 Chile earthquake: inversion for slip distribution from surface deformation. *Geophys. J. Int.* **103**, 589–598 (1990).
23. Angermann, D., Klotz, J. & Reigber, C. Space-geodetic estimation of the Nazca–South America Euler vector. *Earth Planet. Sci. Lett.* **171**, 329–334 (1999).
24. Hyndman, R. D. & Wang, K. Thermal constraints on the zone of major thrust earthquake failure: The Cascadia Subduction Zone. *J. Geophys. Res.* **98**, 2039–2060 (1993).
25. Montgomery, D. R., Balco, G. & Willet, S. D. Climate, tectonics, and the morphology of the Andes. *Geology* **29**, 579–582 (2001).
26. Kulm, L. D., Schweller, W. J. & Masias, A. in *Island Arcs, Deep Sea Trenches and Back-Arc Basins* (eds Talwani, M. & Pitman, W. C. III) 285–301 (Maurice Ewing Ser. 1, American Geophysical Union, 1977).
27. Thornburg, T. & Kulm, L. D. Sedimentation in the Chile Trench: Depositional morphologies, lithofacies, and stratigraphy. *Bull. Geol. Soc. Am.* **98**, 33–52 (1987).
28. Bangs, N. L. & Cande, S. C. Episodic development of a convergent margin inferred from structures and processes along the southern Chile margin. *Tectonics* **16**, 489–503 (1997).
29. Giese, P., Reutter, K. J. & Scheuber, E. *Welche Prozesse haben die Entstehung der extremen Dimensionen der zentralen Anden verursacht? Deformationsprozesse in den Anden* 15–34 (Sonderforschungsbereich 267, Freie Univ. Berlin/Tech. Univ. Berlin/GeoForschungsZentrum Potsdam/Univ. Potsdam, Berlin, 1998).
30. von Huene, R. & Scholl, D. W. Observations at convergent margins concerning sediment subduction, subduction erosion, and the growth of continental crust. *Rev. Geophys.* **29**, 279–316 (1991).
31. von Huene, R. & Ranero, C. R. Subduction erosion and basal friction along the sediment-starved convergent margin of Antofagasta, Chile. *J. Geophys. Res.* **108**, 2079, doi:10.1029/2001JB001569 (2003).
32. Ruff, L. J. Do trench sediments affect great earthquake occurrence in subduction zones? *Pure Appl. Geophys.* **129**, 263–282 (1989).
33. Gutscher, M. A. & Peacock, S. M. Thermal models of flat subduction and the rupture zone of great subduction earthquakes. *J. Geophys. Res.* **108**(B1), 2009, doi:10.1029/2001JB000787 (2003).
34. Clapperton, C. *Quaternary Geology and Geomorphology of South America* (Elsevier, Amsterdam, 1993).
35. Schwertfeger, W. (ed.) *Climates of Central and South America* 1–532 (Elsevier, New York, 1976).
36. Pardo-Casas, F. & Molnar, P. Relative motion of the Nazca (Farallon) and South American plates since Late Cretaceous time. *Tectonics* **6**, 233–248 (1987).
37. Somoza, R. Updated Nazca (Farallon)–South America relative motions during the last 40 My: implications for mountain building in the central Andean region. *J. Soc. Am. Earth Sci.* **11**, 211–215 (1998).
38. Watts, A., Lamb, S., Fairhead, J. & Dewey, J. F. Lithospheric flexure and bending of the central Andes. *Earth Planet. Sci. Lett.* **134**, 9–21 (1995).
39. Hindle, D. *et al.* Consistent geologic and geodetic displacements during Andean orogenesis. *Geophys. Res. Lett.* **29**, 3757–3761 (2002).
40. Zachos, J., Pagani, N., Sloan, L., Thomas, E. & Billups, K. Trends, rhythms, and aberrations in the global climate 65 Ma to present. *Science* **292**, 686–693 (2001).
41. Hartley, A. J. Andean uplift and climate change. *J. Geol. Soc. Lond.* **160**, 7–10 (2003).
42. Strub, P. T. *The Sea–Global Coastal Ocean, Regional Studies and Syntheses*. (eds Robinson, A. R. & Brink, K. H.) 273–315 (Wiley, New York, 1998).
43. Keller, G. *et al.* The Cretaceous/Tertiary boundary event in Ecuador: reduced biotic effects due to the eastern boundary current setting. *Mar. Micropaleontol.* **31**, 97–133 (1997).
44. Rind, D. Latitudinal temperature gradients and climate change. *J. Geophys. Res.* **103**(D6), 5943–5971 (1998).
45. Gubbels, T. L., Isacks, B. L. & Farrar, E. High-level surfaces, plateau uplift, and foreland basin development, Bolivian central Andes. *Geology* **21**, 695–698 (1993).
46. Lamb, S. H. Vertical axis rotation in the Bolivian orocline, South America, 1: Paleomagnetic analysis of Cretaceous and Cenozoic rocks. *J. Geophys. Res.* **106**, 26633–26653 (2001a).
47. Gregory-Wodzicki, K. M. Uplift history of the Central and Northern Andes: A review. *Bull. Geol. Soc. Am.* **112**, 1091–1105 (2000).
48. Lamb, S. H., Hoke, L., Kennan, L. & Dewey, J. in *Orogeny Through Time* (eds Burg, J.-P. & Ford, M.) 237–264 (Geol. Soc. Spec. Publ., Geological Society of London, UK, 1997).
49. Horton, B. K., Hampton, B. A., LaReau, B. N. & Baldellon, E. Tertiary provenance history of the northern and central Altiplano (central Andes, Bolivia): A detrital record of plateau margin tectonics. *J. Sedim. Res.* **72**, 711–726 (2002).
50. Crouch, S. T. Rifts and swells: Geophysical constraints on causality. *Tectonophysics* **94**, 23–37 (1983).
51. Smith, A. G. in *Thrust and Nappe Tectonics* (eds McClay, K. R. & Price, N. J.) 111–124 (Geol. Soc. Spec. Publ. 9, Geological Society of London, UK, 1981).
52. Beck, S. *et al.* Crustal thickness variations in the central Andes. *Geology* **24**, 407–410 (1996).

Acknowledgements This work was supported by grants from the European Union, Natural Environment Research Council, and Royal Society (S.H.L.) and a visiting Leverhulme professorship (P.D.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.H.L. (simon.lamb@earth.ox.ac.uk).

Genome-scale approaches to resolving incongruence in molecular phylogenies

Antonis Rokas*, Barry L. Williams*, Nicole King & Sean B. Carroll

Howard Hughes Medical Institute, Laboratory of Molecular Biology, R. M. Bock Laboratories, University of Wisconsin-Madison, 1525 Linden Drive, Madison, Wisconsin 53706, USA

* These authors contributed equally to this work

One of the most pervasive challenges in molecular phylogenetics is the incongruence between phylogenies obtained using different data sets, such as individual genes. To systematically investigate the degree of incongruence, and potential methods for resolving it, we screened the genome sequences of eight yeast species and selected 106 widely distributed orthologous genes for phylogenetic analyses, singly and by concatenation. Our results suggest that data sets consisting of single or a small number of concatenated genes have a significant probability of supporting conflicting topologies. By contrast, analyses of the entire data set of concatenated genes yielded a single, fully resolved species tree with maximum support. Comparable results were obtained with a concatenation of a minimum of 20 genes; substantially more genes than commonly used but a small fraction of any genome. These results have important implications for resolving branches of the tree of life.

Understanding the historical relationships between living organisms has been one of the principal goals of evolutionary research. Molecular phylogenetic data are instrumental in research on the history of life^{1–3}, the polarity of phenotypic and developmental evolution⁴, and on the diversity of living organisms⁵. Despite tremendous progress in recent years, phylogenetic reconstruction involves many challenges that create uncertainty with respect to the true historical associations of the taxa analysed. One of the most notable difficulties is the widespread occurrence of incongruence between alternative phylogenies generated from single-gene data sets. Incongruence occurs at all taxonomic levels, from phylogenies of closely related species^{6,7} to relationships between major classes^{8,9} or phyla and higher taxonomic groups^{10–12}.

Both analytical and biological factors may cause incongruence^{10,13}. Analytical factors affecting phylogenetic reconstruction include the choice of optimality criterion¹⁴, limited data availability^{15,16}, taxon sampling¹⁷ and specific assumptions in the modelling of sequence evolution¹⁸. Biological processes such as the action of natural selection or genetic drift^{19–22} may cause the history of the genes under analysis to obscure the history of the taxa. The large number of potential explanations for the presence of incongruence in molecular phylogenetic analyses makes decisions on how to handle conflict in larger sets of molecular data difficult²³. For example, two genes with different evolutionary histories (for example, owing to hybridization or horizontal transfer) for a particular taxonomic group will by definition be incongruent while still depicting true histories²⁰. Data sets composed of genes showing heterogeneity in mode of sequence evolution may also compound bias rather than resolve the true history²⁴. Furthermore, because current tests are not always reliable^{25,26}, it has been difficult to estimate incongruence. To overcome the effect of analytical and biological factors by increasing the signal-to-noise ratio, many researchers have attempted to address difficult phylogenetic questions by analysis of concatenated data sets^{1,27–29}. However, phylogenetic analyses of different sets of concatenated genes do not always converge on the same tree^{8,9}, and some studies have yielded results at odds with widely accepted phylogenies³⁰. Although theory suggests that a number of factors (such as gene number, sequence length, optimality criterion and rate of evolution) may influence phylogenetic reconstruction^{14,15,31}, the effect of these factors has not been systematically explored with large data sets derived from biological sequences. Recent progress in the genomics of the

yeast genus *Saccharomyces*^{32–34} has presented an unprecedented opportunity to evaluate these issues in eukaryotic phylogenetics.

Screen for and phylogenetic analysis of orthologous genes

Genome sequence data have been obtained for seven *Saccharomyces* species (*S. cerevisiae*, *S. paradoxus*, *S. mikatae*, *S. kudriavzevii*, *S. bayanus*, *S. castellii* and *S. kluyveri*)^{32–34} as well as for the outgroup fungus *Candida albicans*. The genomes of all eight species were screened for orthologous genes to serve as phylogenetic markers (see Methods). Previous work^{35–37} has suggested the occurrence of a genome duplication in the evolutionary history of the *Saccharomyces* yeasts, so our gene selection criteria relied on synteny to establish orthology (see Methods). We retained 106 genes, which are distributed throughout the *S. cerevisiae* genome on all 16 chromosomes and comprise a total length of 127,026 nucleotides (42,342 amino acids), corresponding to roughly 1% of the genomic sequence and 2% of the predicted genes. Phylogenetic reconstructions were performed in three ways: maximum likelihood (ML) analysis of the nucleotide data, and maximum parsimony (MP) analysis of both the nucleotide and the amino acid data. Because the study comprised eight taxa, we used a search strategy in each analysis that was guaranteed to find the most parsimonious or most likely tree with respect to each gene (see Methods).

Single-gene phylogenies reveal extensive incongruence

Analyses of the 106 genes resulted in more than 20 alternative ML or MP trees (see Methods). To assess the degree of support for each of these trees, 100 bootstrap replicates were generated and summarized as 50% majority-rule consensus trees. Several of the many strongly supported (defined here as a bootstrap value >70%) alternative phylogenies are shown in Fig. 1a–f. For example, some genes recovered strong support for various alternative placements of species within the *sensu stricto* group (*S. cerevisiae*, *S. paradoxus*, *S. mikatae*, *S. kudriavzevii* and *S. bayanus*; Fig. 1a, c–e), whereas other genes strongly supported alternative placements of *S. castellii* and *S. kluyveri* relative to the outgroup (Fig. 1a, b, f).

The 106 genes analysed were selected without consideration of their function, and their phylogenetic performance has not been evaluated previously. Genes more commonly used in molecular systematics are often chosen for both historical (for example, ease of amplification, previous analyses of related taxa) and analytical (for example, copy number, rate of evolution) reasons³⁸. It is possible

that commonly used genes may provide better resolution of the same set of taxa. We repeated ML and MP analyses of nucleotide sequence on a sample of six commonly used genes (actin, hsp70, β -tubulin, RNA polymerase II, elongation factor 1- α and 18S rDNA), and also obtained alternative and often strongly supported phylogenies (Fig. 1g–l). Specifically, actin and RNA polymerase II (Fig. 1g, j) each support the same tree as in Fig. 1a, whereas hsp70 (Fig. 1h) supports the topology in Fig. 1b. β -Tubulin (Fig. 1i), 18S rDNA (Fig. 1l) and elongation factor 1- α (Fig. 1k) each support, albeit weakly at some branches, trees not represented in Fig. 1a–f. Therefore, analyses of both commonly used genes and the set of 106 orthologous genes provide strong support for several alternative phylogenies and fail to indicate which gene tree(s) might represent the actual species tree.

The alternative phylogenies could have resulted from a number of different scenarios: (1) most genes could have weakly supported most phylogenies and strongly supported only a few alternative trees; (2) most genes could have strongly supported one phylogeny and a few genes strongly supported only a small number of alternatives; (3) there could have been some combination of these scenarios so that each branch among alternative phylogenies had either weak or strong support depending on the gene. To distinguish between these possibilities, we identified all of the branches recovered during the single-gene analyses, and recorded each bootstrap value with respect to the gene and method of analysis (see Supplementary Information). Eight branches were shared by all three analyses with multiple instances of bootstrap values >50%; hence, we focus on these eight branches in our analyses.

For each branch, we observed a full range of bootstrap values (Fig. 2c). Only two branches (branches 1 and 4 in Fig. 2) were supported with high bootstrap values by a majority of genes. The remaining six branches were supported by bootstrap values that range evenly from 0 to 100 (Fig. 2c). Two summary statistics of these distributions, the mean bootstrap value and the percentage of genes supporting each branch (Fig. 2c), also indicate weak support overall

for each of these six branches (Fig. 2c). Notably, bootstrap values for branches 3 versus 6 and among branches 5, 7 and 8 exhibit significant and strong negative correlations so that a low bootstrap value for one branch corresponds to a high bootstrap value for an alternative branch (Pearson correlation coefficients and *P*-values are: -0.684 and <0.001 , -0.734 and <0.001 , and -0.607 and <0.001 for branches 3 versus 6, 5 versus 7, and 5 versus 8, respectively; values are for ML analysis; similar results were obtained from the remaining two MP analyses (data not shown)). Although the values of some other pairs of branches were also correlated, these were the only pairs of branches exhibiting a very strong negative correlation across all three analyses. This negative correlation is indicative of the pervasive direct conflict among genes with respect to these branches. In summary, the support for a given branch was strongly dependent on the gene analysed.

Although the trees recovered from single-gene analyses are incongruent in that they exhibit topological differences, the degree of conflict among trees could be relatively minor. We examined the degree of incongruence between the trees recovered by determining the number of taxa that would need to be removed in order to make two trees congruent (see Methods). These values were computed for all possible pairwise comparisons among the 106 50% majority-rule consensus trees generated within each of the three analyses. The distribution of values shows that only a small proportion of all pairwise comparisons are in total agreement ($<15\%$), whereas most of the trees ($>50\%$) require that at least two of the eight taxa are removed before they become congruent, regardless of the method of analysis (Fig. 3). These results were consistent with other metrics of incongruence (see Methods and Supplementary Information) and indicate extensive incongruence among the 106 data sets.

Factors potentially influencing incongruence between trees

We mitigated some potential sources of incongruence by analysing a genome-wide sample of orthologous genes. There are a number of

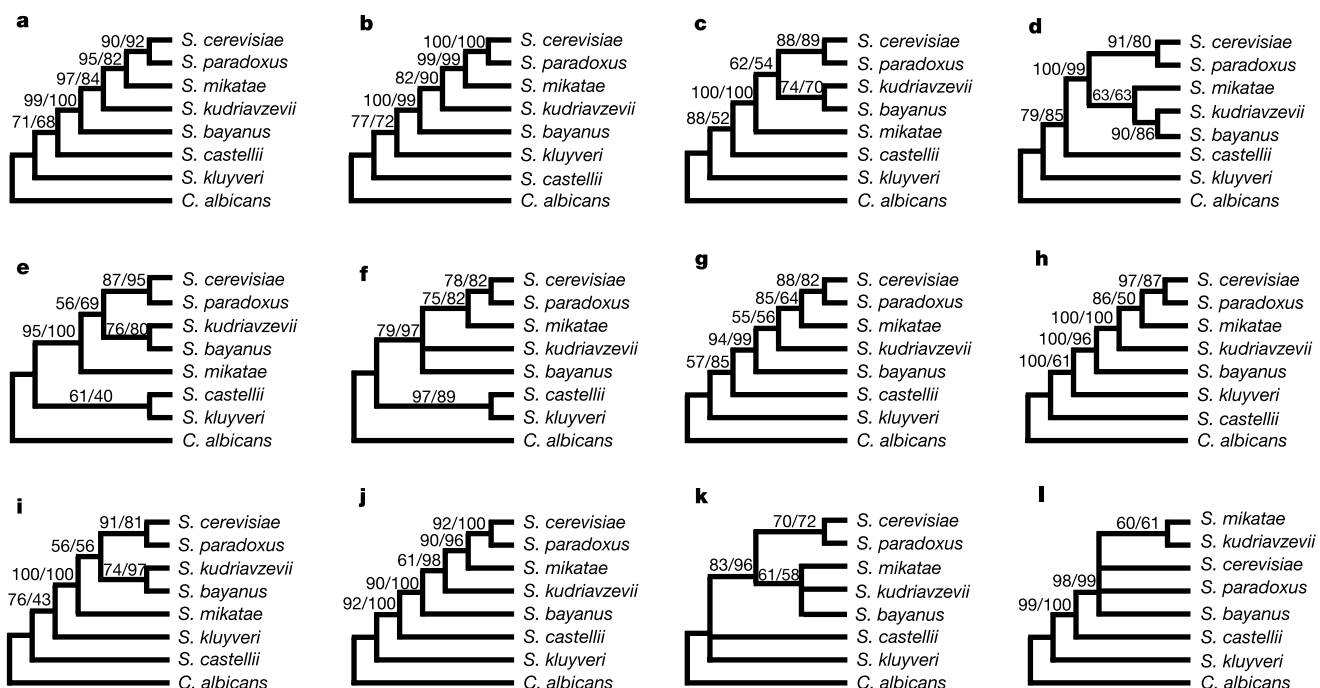


Figure 1 Single-gene data sets generate multiple, robustly supported alternative topologies. Representative alternative trees recovered from analyses of nucleotide data of 106 selected single genes and six commonly used genes are shown. The trees are the 50% majority-rule consensus trees from the genes YBL091C (a), YDL031W (b),

YER005W (c), YGL001C (d), YNL155W (e) and YOL097C (f), as well as those from the commonly used genes actin (g), hsp70 (h), β -tubulin (i), RNA polymerase II (j) elongation factor 1- α (k) and 18S rDNA (l). Numbers above branches indicate bootstrap values (ML on nucleotides/MP on nucleotides).

additional analytical and biological factors—such as outgroup choice, number of variable sites and rate of evolution—that may lead to incongruence between single-gene phylogenies^{10,13}. To test whether the outgroup accounted for the incongruence between phylogenies, we repeated all of the analyses without the outgroup *C. albicans*. We found no change in the distribution of bootstrap values (correlations among pairwise comparisons of each distribution for the remaining branches were significant with $P < 0.05$) or in the degree of incongruence between the remaining branches (Supplementary Information). We also examined whether support for each branch was explained by the number of variable sites, number of parsimony-informative sites, gene size, rate of evolution, nucleotide composition, base compositional bias, genome location, or gene ontology (Table 1; see also Supplementary Information). Number of variable sites, number of parsimony-informative sites and gene size were significantly correlated with bootstrap values for some branches, although they accounted for only a small amount of the total variation in each case (Table 1; see also Supplementary Information). With a single exception (branch 4 was correlated with the rate of evolution for the ML analysis; Table 1), none of the remaining variables was correlated with bootstrap values for any branch (Table 1; see also Supplementary Information). In summary, there were no identifiable parameters that could systematically account for or predict the performance of single genes.

Concatenation of single genes yields a single tree

Although we do not know the cause(s) of incongruence between single-gene phylogenies, the critical question is how the pervasive incongruence between single trees might be overcome to arrive at the actual species tree. Although many potential options exist, we explored the effect of concatenating single genes into one large data set^{1,27,39}. Remarkably, all three methods of analysis of the concatenated sequences yielded a single tree with 100% bootstrap values at every branch (Fig. 4). Furthermore, all alternative topologies generated among the single-gene analyses were rejected (Templeton test, $P < 0.001$ for each of three analyses). Thus, even though the individual genes examined supported alternative trees, the concatenated data exclusively supported a single tree. This level of support for a single tree with five internal branches is, to our knowledge, unprecedented; we conclude that it accurately represents the historical relationships of these eight yeast taxa and will be referred to hereafter as their species tree. The maximum support for a single topology regardless of method of analysis is strongly suggestive of the power of large data sets in overcoming the incongruence present in single-gene analyses.

How much data are sufficient to recover the species tree?

The concatenated data recovered a tree with maximum support on all branches, despite divergent levels of support for each branch among single-gene analyses. This raises the question: at what size

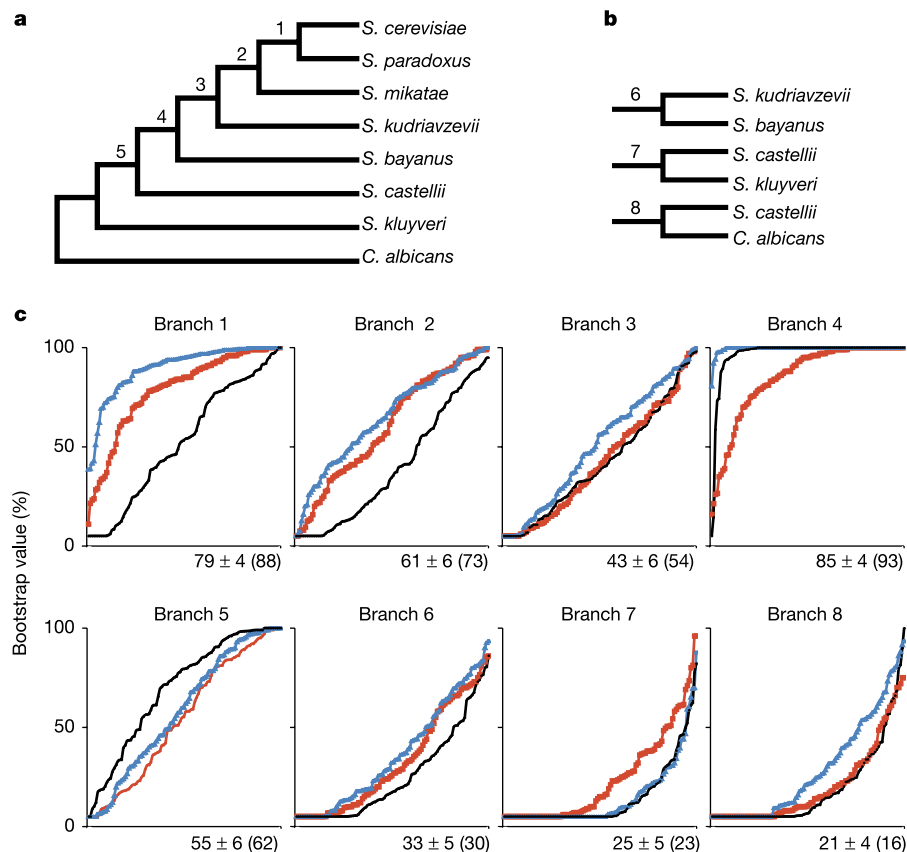


Figure 2 The distribution of bootstrap values for the eight prevalent branches recovered from 106 single-gene analyses highlights the pervasive conflict among single-gene analyses. **a**, Majority-rule consensus tree of the 106 ML trees derived from single-gene analyses. Across all analyses, there were eight commonly observed branches; the five branches in the consensus tree (numbers 1–5; **a**) and the three branches (numbers 6–8) shown in **b**, **c**. For each of the eight branches, the ranked distribution of per cent bootstrap values recovered from the three analyses of 106 genes is shown. Results from ML (blue

and MP (red) analyses of nucleotide data sets, and MP analyses of amino acid data sets (black), are shown. For each branch, the mean bootstrap value and 95% confidence intervals from the ML analyses and the percentage of ML trees supporting this branch (in parentheses) are indicated below each graph. Although the ranked distributions of bootstrap values from the three analyses are remarkably similar for most branches, on a gene-by-gene basis there is no tight correspondence between bootstrap values from ML and MP analyses (see Supplementary Information).

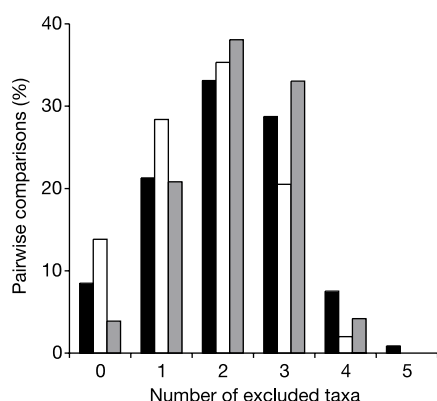


Figure 3 Extensive incongruence between trees derived from the 106 individual-gene data sets. Pairwise comparisons between 50% majority-rule consensus trees from 106 single-gene ML analyses of nucleotide data (black bars), MP analyses of nucleotide data (white bars), and MP analyses of amino acid data (grey bars) were categorized on the basis of the minimum number of taxa that need to be removed for two trees to reach congruence (x axis). For each of the analyses, the majority of pairwise comparisons require the removal of two or more taxa before congruence is attained. Similar results are obtained when the ML and MP trees from the analyses of the 106 genes are used (data not shown).

did the data set arrive at the species tree? To estimate the minimum number of genes required to support the species tree (Fig. 4), we randomly re-sampled and concatenated variable numbers of genes from the 106-gene data set (Fig. 5; see also Supplementary Information). Although the number of genes required to achieve a mean bootstrap value $>70\%$ across all branches of the species tree was only three concatenated genes, the variance in this estimate was high (Fig. 5, right panels). The results show that in order to achieve a mean bootstrap value of at least 70% with a 95% confidence interval, eight concatenated genes were required. For a mean bootstrap value of at least 95% with a confidence interval of 95%, the number of concatenated genes rose to 20. These results show that the number of genes sufficient to support all branches of the species tree ranged from a minimum of 8 to 20 (Fig. 5, right panels), depending on the threshold of statistical support required.

Because we observed a high variance in bootstrap values from small numbers of concatenated genes, we sought to explore the underlying source of this variance. It has been suggested that nucleotides within a given gene do not evolve independently, thus potentially influencing the phylogenetic accuracy of single genes¹⁶. To test this hypothesis, we used a variable length bootstrap procedure in which a subset of orthologous nucleotides was randomly

re-sampled from the total data set with replacement. The results show that with only 3,000 nucleotides, all five branches were supported with a mean bootstrap value $>70\%$ and a confidence interval of 95% (Fig. 5, left panels; see also Supplementary Information). With 8,000 nucleotides, the mean bootstrap value rose to $>95\%$ with a confidence interval of 95%. The average size of a gene in our data set was 1,198 base pairs, so 3,000 randomly selected nucleotides correspond to less than three concatenated genes. Importantly, random nucleotides had much lower variance in bootstrap values when compared with corresponding numbers of concatenated genes. For example, with 3,000 nucleotides the mean bootstrap value ($\pm 95\%$ confidence interval) was 81.03 ± 1.08 and 91.06 ± 0.83 for branches 3 and 5, respectively, whereas with three concatenated genes the corresponding bootstrap values were 74.36 ± 11.69 and 70.98 ± 18.48 (Fig. 5). The lower bootstrap value and much higher variance for concatenated genes relative to randomly re-sampled nucleotides is consistent with the hypothesis that nucleotides within genes have not evolved independently^{16,30,40}.

The results demonstrate that concatenation of a sufficient number of randomly selected genes overwhelms conflicting signals present in different genes. It is important to determine whether a consistent bias present in a subset of genes is also overcome by concatenation. To test this possibility, we concatenated genes with bootstrap values $>50\%$ for each of the alternative branches 6, 7 and 8 (Fig. 2). Each analysis yielded a majority-rule consensus tree with $>90\%$ bootstrap values at most branches, but none of the trees was congruent with that recovered from concatenation of the full data set (Fig. 6). For example, the six genes that most strongly supported *S. castellii* and *S. kluyveri* as sister taxa recovered these species as sister taxa with 100% bootstrap values when concatenated (Fig. 6b). All three trees in Fig. 6 also rejected all alternative topologies generated by single-gene and concatenated analyses (Templeton test, $P < 0.001$ for each of the three trees in Fig. 6). Hence, when biased genes present in the data set were concatenated, strong and misleading support for an alternative species tree was obtained. Previous studies have shown that the concatenation of genes that share some bias (for example, mitochondrial genes) can produce strong support for the incorrect phylogeny³⁰. Thus, concatenation of a large number of unlinked genes is clearly the superior strategy.

Implications for resolution of phylogenies

Our results show that there is widespread incongruence between phylogenies recovered from individual genes. Therefore reliance on single or a small number of genes has a significant probability of supporting incorrect relationships for the eight yeast taxa. Perhaps surprisingly, none of the factors known or predicted to cause phylogenetic error⁴¹ could systematically account for the observed

Table 1 Regressions of bootstrap values on analytical factors

Branch*	Factor							
	Variable sites		Informative sites†		Branch lengths‡		(G + C)%	
	r^2	P-value†	r^2	P-value	r^2	P-value	r^2	P-value
1	0.135	<0.001	0.091	0.002	0.040	0.041	<0.001	0.873
2	0.204	<0.001	0.120	<0.001	0.014	0.228	0.005	0.481
3	0.150	<0.001	0.115	<0.001	0.016	0.197	0.009	0.321
4	0.135	<0.001	0.042	0.035	0.110	0.001	0.022	0.134
5	0.036	0.051	0.073	0.005	0.016	0.200	0.004	0.513
6	0.052	0.018	0.099	0.001	0.002	0.640	0.006	0.432
7	0.072	0.005	0.122	<0.001	0.006	0.413	0.001	0.735
8	0.007	0.392	0.003	0.557	0.001	0.775	<0.001	0.869

*Branch numbers correspond to Fig. 2 and are provided for ML analyses only (see Supplementary Information for the remainder of the analyses).

†Significant values are <0.006 after Bonferroni correction for tests between eight branches.

‡Shown for MP analyses of the nucleotide data only.

§Branch lengths estimated using ML.

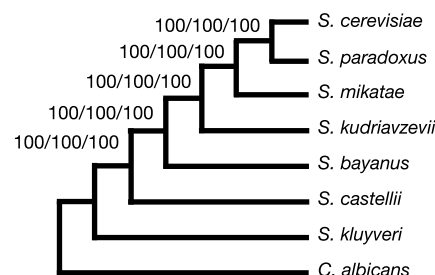


Figure 4 Phylogenetic analyses of the concatenated data set composed of 106 genes yield maximum support for a single tree, irrespective of method and type of character evaluated. Numbers above branches indicate bootstrap values (ML on nucleotides/MP on nucleotides/MP on amino acids).

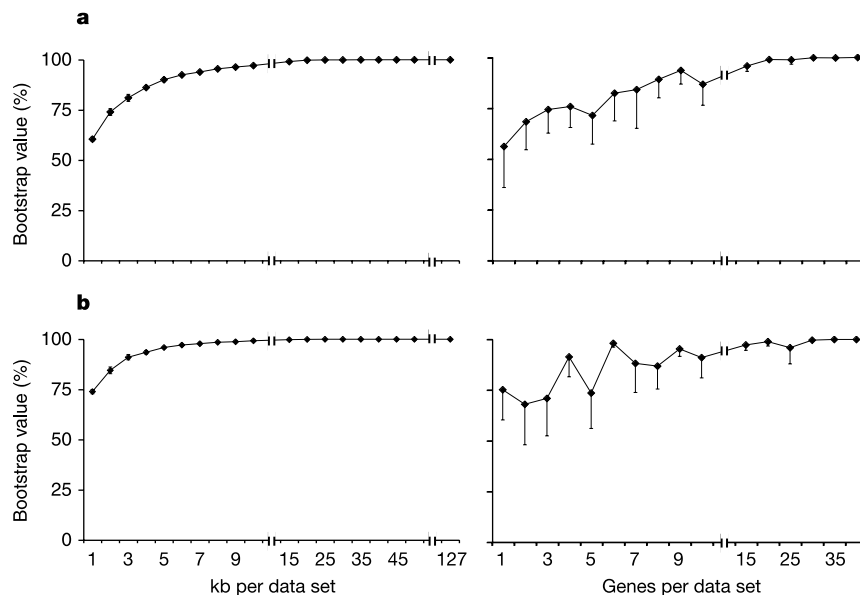


Figure 5 A minimum of 20 genes is required to recover >95% bootstrap values for each branch of the species tree. **a, b**, The bootstrap values for branches 3 (**a**) and 5 (**b**) were constructed from the concatenation of randomly re-sampled orthologous nucleotides (left) or random subsets of genes (right) (for the remaining branches see Supplementary Information). Each data point represents the mean and minus one confidence interval of ten replicates (confidence interval values for left panels are not clearly visible due to almost complete overlap with the mean). The species tree is recovered with robust support (>95% bootstrap values in all branches at 95% confidence interval) by analyses

of a minimum of 20 concatenated genes. This same result is obtained using only 8,000 randomly selected orthologous nucleotides, indicating that nucleotides within genes evolve non-independently. The variation associated with the confidence intervals in **b** (for example, for branch 5 the data point for 25 genes has a larger confidence interval than that for 20 genes) is due to the small number of replicates. Broken lines in each graph represent a change in the linear values along the x axis. All analyses were performed using MP.

incongruence, suggesting that there may be no good predictor of the phylogenetic informativeness of genes. However, regardless of the source of incongruence, concatenation of a sufficient number of unlinked genes (≥ 20 genes in this study) yields the species tree with remarkable support.

These findings have important implications for many current practices in molecular phylogenetics. A strict interpretation of our data suggests that analyses based on single or a small number of genes provide insufficient evidence for establishing or refuting phylogenetic hypotheses. Furthermore, concatenations of small numbers of genes that provide support for specific branches leads to amplification of support for that branch, even if the branch is in favour of an incorrect topology. For example, a recent phylogenetic study of a concatenation of eight commonly used genes for 75 species belonging to the ‘*Saccharomyces* complex’ found a bootstrap value of 69% in support of a sister group relationship between *S. paradoxus* and *S. mikatae*⁴², a finding in sharp contrast to our results. This may be accounted for by both the number and non-

independence of the set of genes examined. The unreliability of single-gene data sets (or data sets composed of linked genes, such as genes from the mitochondrial genome) stems from the fact that each gene is shaped by a unique set of functional constraints through evolution. Phylogenetic algorithms are sensitive to such constraints³⁰ and we show that with genome-wide sampling of independently evolving genes such problems can be avoided. Of course, the possibility exists that one or a few ‘lucky’ genes³⁹ may correctly reconstruct the phylogeny of a particular taxonomic group but, in the absence of other data, the confidence in the conclusion will be weak (Fig. 5).

It is only through analyses of a larger amount of sequence data that confidence in the proposed phylogenetic reconstruction can be obtained. In this phylogenetic analysis of eight yeast taxa, concatenated data sets of 20 genes are sufficient to provide very strong (>95%) support for the species tree. It is possible that the eight yeast taxa we have analysed represent a very difficult phylogenetic case, atypical of the situations found in other groups. However, the

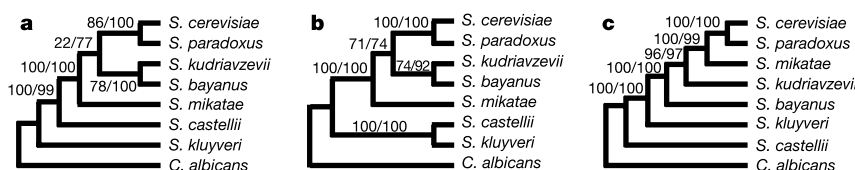


Figure 6 Concatenation of genes supporting alternative branches leads to further amplification of bias. **a–c**, Concatenation of the ten single genes showing the highest bootstrap values for a sister-group relationship between *S. kudriavzevii* and *S. bayanus* (**a**), the six data sets showing bootstrap values >50% for a sister-group relationship between *S. castellii* and *S. kluyveri* (**b**), or the ten data sets showing the highest bootstrap values for *S. castellii* being the most distant member of the ingroup (**c**) generates tree topologies that are inconsistent with the species tree recovered from concatenation of 106

genes (Fig. 4). These results may be explained either by a systematic bias imposed by analytical factors (for example, variable rates of nucleotide site evolution) misleading phylogenetic algorithms to artefactually group certain taxa, or by the action of biological factors (for example, certain genes having different genealogical histories from the species owing to hybridization). Numbers above branches indicate bootstrap values (ML on nucleotides/MP on nucleotides).

widespread occurrence of incongruence at all taxonomic levels argues strongly against such a view^{6,8–10,21}. Rather, we believe that this group is a representative model for key issues that researchers in phylogenetics are confronting.

Of course, in other cases the amount of sequence information needed to resolve specific relationships will be dependent on the particular phylogenetic history under examination. For example, branches depicting speciation events separated by long time intervals may be resolved with a small amount of data (perhaps branches 1 and 4 in our analyses), whereas branches depicting speciation events separated by shorter intervals may be much harder to resolve. In addition, phylogenetic reconstruction may be complicated by factors such as taxon sampling¹⁷, variable rates of nucleotide site evolution⁴³, hybridization⁴⁴, horizontal transfer⁴⁵ and lineage sorting of ancestral polymorphisms²¹. These factors may have an as yet unpredictable effect on the amount of sequence data required for accurate phylogenetic reconstruction of other taxonomic groups or of larger numbers of taxa. Larger concatenated data sets have been very powerful and persuasive in testing specific relationships (such as monophyly of amoebae³⁹ and rejection of Ecdysozoa⁴⁶). The results of our study suggest that use of genome-wide data sets may provide unprecedented power not only in testing specific phylogenetic hypotheses but also in precise reconstruction of the historical associations of all the taxa analysed. □

Methods

Data set collection

Genes spaced approximately every 40 kilobases (kb) in the *S. cerevisiae* genome were examined and retained for analysis if they met the following criteria: (1) had homologous sequence in each of the eight species; (2) had at least two homologous flanking syntenic genes; and (3) could be aligned over most of the protein. The species of the *sensu strictu* group are thought to have undergone a whole-genome duplication^{35–37}, so determination of orthology was based on synteny. For *S. paradoxus*, *S. mikatae* and *S. bayanus*, synteny relative to *S. cerevisiae* was determined by visual inspection using the synteny viewer option in the *Saccharomyces* Genome Database (<http://www.yeastgenome.org/>). For *S. kudriavzevii*, *S. castellii* and *S. kluyveri*, synteny was determined by tblastx comparisons of the *S. cerevisiae* gene of interest, plus each of the two flanking genes. Genes were considered as orthologues if the gene of interest and at least one of the two flanking genes had significant tblastx scores (*e*-value <0.001) and were found in close proximity to each other in all genomes. For *C. albicans*, genes were retained if they were the best reciprocal tblastx hits between both *S. cerevisiae* and *C. albicans*. The genome databases for *S. kudriavzevii*, *S. castellii* and *S. kluyveri* contain draft quality sequence providing 2–3 times coverage over 85–95% of each genome³⁴. Therefore, some cases of segmental duplications may have been missed; such cases would be rare and have had little, if any, influence on our results. The selection of commonly used genes (actin, hsp70, β -tubulin, RNA polymerase II, elongation factor 1- α and 18S rDNA) involved searching each genome for the single best tblastx score (in cases of paralogy, the hit with the highest tblastx score was considered as the orthologue). Many other commonly used genes were not analysed because a copy was absent in at least one of the genome databases. Data sets are available from the authors on request.

Phylogenetic analysis

Individual genes were aligned by codon using ClustalW⁴⁷ as implemented in Bioedit version 5.0.9 (ref. 48). All gene alignments were manually edited to exclude indels and areas of uncertain alignment from further analysis. After this step, on average 76% of the sequence of each gene was retained. All phylogenetic analyses were performed using PAUP* version 4.0b10 (ref. 49). Each nucleotide data set was analysed under the optimality criteria of maximum likelihood (ML) and maximum parsimony (MP), whereas the corresponding amino acid data sets were analysed using only MP. In both ML and MP analyses, tree space was searched using the branch-and-bound algorithm, which guarantees to find the optimal tree(s). Tree reliability under both optimality criteria was assessed using non-parametric bootstrap re-sampling of 100 replicates; for a given data set, bootstrap re-sampling for both ML and MP analyses started from the same seed number to eliminate variance in the re-sampling procedure between the two optimality criteria. Under MP all characters were equally weighted whereas under ML the model of sequence evolution was optimized using likelihood ratio tests as implemented in Modeltest version 3.06 (ref. 50). Analyses of concatenated data sets were performed using MP (1,000 bootstrap replicates) and, in a subset of cases, ML (100 bootstrap replicates). Random re-sampling of individual nucleotide sites from the complete 127-kb data set was performed using the variable-length-bootstrap option in PAUP*⁴⁹ and random gene re-sampling using a random number generator. Incongruence was tested using Templeton's significantly less parsimonious test, and tree distances for all pairwise comparisons among the 106 trees were calculated using two metrics (the agreement subtree metric and the symmetric-difference metric) as implemented in PAUP*⁴⁹. Data sets and trees are available from the authors on request.

Received 4 June; accepted 15 September 2003; doi:10.1038/nature02053.

- Baldauf, S. L., Roger, A. J., Wenk-Siefert, I. & Doolittle, W. F. A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science* **290**, 972–977 (2000).
- Brown, J. R., Douady, C. J., Italia, M. J., Marshall, W. E. & Stanhope, M. J. Universal trees based on large combined protein sequence data sets. *Nature Genet.* **28**, 281–285 (2001).
- Aguinaldo, A. M. A. *et al.* Evidence for a clade of nematodes, arthropods and other moulting animals. *Nature* **387**, 489–493 (1997).
- Knoll, A. H. & Carroll, S. B. Early animal evolution: emerging views from comparative biology and geology. *Science* **284**, 2129–2137 (1999).
- Pace, N. R. A molecular view of microbial diversity and the biosphere. *Science* **276**, 734–740 (1997).
- Kopp, A. & True, J. R. Phylogeny of the Oriental *Drosophila melanogaster* species group: a multilocus reconstruction. *Syst. Biol.* **51**, 786–805 (2002).
- Mason-Gamer, R. J. & Kellogg, E. A. Testing for phylogenetic conflict among molecular data sets in the tribe Triticeae (Gramineae). *Syst. Biol.* **45**, 522–543 (1996).
- Giribet, G., Edgecombe, G. D. & Wheeler, W. C. Arthropod phylogeny based on eight molecular loci and morphology. *Nature* **413**, 157–161 (2001).
- Hwang, U. W., Friedrich, M., Tautz, D., Park, C. J. & Kim, W. Mitochondrial protein phylogeny joins myriapods with chelicerates. *Nature* **413**, 154–157 (2001).
- Rokas, A., King, N., Finnerty, J. & Carroll, S. B. Conflicting phylogenetic signals at the base of the metazoan tree. *Evol. Dev.* **5**, 346–359 (2003).
- Loytynoja, A. & Milinkovitch, M. C. Molecular phylogenetic analyses of the mitochondrial ADP-ATP carriers: the Plantae/Fungi/Metazoa trichotomy revisited. *Proc. Natl Acad. Sci. USA* **98**, 10202–10207 (2001).
- Baldauf, S. L. & Palmer, J. D. Animals and fungi are each other's closest relatives: congruent evidence from multiple proteins. *Proc. Natl Acad. Sci. USA* **90**, 11558–11562 (1993).
- Wendel, J. F. & Doyle, J. J. in *Molecular Systematics of Plants II: DNA Sequencing* (eds Soltis, D. E., Soltis, P. S. & Doyle, J. J.) 265–296 (Kluwer, Boston, Massachusetts, 1998).
- Huelsenbeck, J. P. Performance of phylogenetic methods in simulation. *Syst. Biol.* **44**, 17–48 (1995).
- Philippe, H., Chenuil, A. & Adoutte, A. Can the cambrian explosion be inferred through molecular phylogeny? *Development* (Suppl.) 15–25 (1994).
- Cummings, M. P., Otto, S. P. & Wakeley, J. Sampling properties of DNA sequence data in phylogenetic analysis. *Mol. Biol. Evol.* **12**, 814–822 (1995).
- Graybeal, A. Is it better to add taxa or characters to a difficult phylogenetic problem? *Syst. Biol.* **47**, 9–17 (1998).
- Yang, Z., Goldman, N. & Friday, A. Comparison of models for nucleotide substitution used in maximum-likelihood phylogenetic estimation. *Mol. Biol. Evol.* **11**, 316–324 (1994).
- Martin, A. P. & Burg, T. M. Perils of paralogy: using HSP70 genes for inferring organismal phylogenies. *Syst. Biol.* **51**, 570–587 (2002).
- Maddison, W. P. Gene trees in species trees. *Syst. Biol.* **46**, 523–536 (1997).
- Satta, Y., Klein, J. & Takahata, N. DNA archives and our nearest relative: the trichotomy problem revisited. *Mol. Phylog. Evol.* **14**, 259–275 (2000).
- Rieseberg, L. H., Whittton, J. & Linder, C. R. Molecular marker incongruence in plant hybrid zones and in phylogenetic trees. *Acta Bot. Neerland.* **45**, 243–262 (1996).
- Huelsenbeck, J. P., Bull, J. J. & Cunningham, C. W. Combining data in phylogenetic analysis. *Trends Ecol. Evol.* **11**, 152–158 (1996).
- Bull, J. J., Huelsenbeck, J. P., Cunningham, C. W., Swofford, D. L. & Waddell, P. J. Partitioning and combining data in phylogenetic analysis. *Syst. Biol.* **42**, 384–397 (1993).
- Sullivan, J. Combining data with different distributions of among-site rate variation. *Syst. Biol.* **45**, 375–380 (1996).
- Cunningham, C. W. Can three incongruence tests predict when data should be combined? *Mol. Biol. Evol.* **14**, 733–740 (1997).
- Soltis, P. S., Soltis, D. E. & Chase, M. W. Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. *Nature* **402**, 402–404 (1999).
- Murphy, W. J. *et al.* Molecular phylogenetics and the origins of placental mammals. *Nature* **409**, 614–618 (2001).
- Moreira, D., Le Guyader, H. & Philippe, H. The origin of red algae and the evolution of chloroplasts. *Nature* **405**, 69–72 (2000).
- Naylor, G. J. P. & Brown, W. M. Amphioxus mitochondrial DNA, chordate phylogeny, and the limits of inference based on comparisons of sequences. *Syst. Biol.* **47**, 61–76 (1998).
- Hillis, D. M. Inferring complex phylogenies. *Nature* **383**, 130–131 (1996).
- Kellis, M., Patterson, N., Endrizzi, M., Birren, B. & Lander, E. S. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* **423**, 241–254 (2003).
- Goffeau, A. *et al.* Life with 6000 genes. *Science* **274**, 546 (1996) 563–567.
- Cliften, P. *et al.* Finding functional features in *Saccharomyces* genomes by phylogenetic footprinting. *Science* **301**, 71–76 (2003).
- Wolfe, K. H. & Shields, D. C. Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* **387**, 708–713 (1997).
- Wong, S., Butler, G. & Wolfe, K. H. Gene order evolution and paleopolyploidy in hemiascomycete yeasts. *Proc. Natl Acad. Sci. USA* **99**, 9272–9277 (2002).
- Langkjaer, R. B., Cliften, P. F., Johnston, M. & Piskur, J. Yeast genome duplication was followed by asynchronous differentiation of duplicated genes. *Nature* **421**, 848–852 (2003).
- Hillis, D. M., Moritz, C. & Mable, B. K. (eds) *Molecular Systematics* (Sinauer, Sunderland, Massachusetts, 1996).
- Baptiste, E. *et al.* The analysis of 100 genes supports the grouping of three highly divergent amoebae: *Dictyostelium*, *Entamoeba*, and *Mastigamoeba*. *Proc. Natl Acad. Sci. USA* **99**, 1414–1419 (2002).
- Averof, M., Rokas, A., Wolfe, K. H. & Sharp, P. M. Evidence for a high frequency of simultaneous double-nucleotide substitutions. *Science* **287**, 1283–1286 (2000).
- Sanderson, M. J. & Shaffer, H. B. Troubleshooting molecular phylogenetic analyses. *Annu. Rev. Ecol. Syst.* **33**, 49–72 (2002).
- Kurtzman, C. P. & Robnett, C. J. Phylogenetic relationships among yeasts of the 'Saccharomyces complex' determined from multigene sequence analyses. *FEMS Yeast Res.* **3**, 417–432 (2003).
- Lopez, P., Casane, D. & Philippe, H. Heterotachy, an important process of protein evolution. *Mol. Biol. Evol.* **19**, 1–7 (2002).

44. Naumov, G. I., James, S. A., Naumova, E. S., Louis, E. J. & Roberts, I. N. Three new species in the *Saccharomyces sensu stricto* complex: *Saccharomyces cariocanus*, *Saccharomyces kudriavzevii* and *Saccharomyces mikatae*. *Int. J. Syst. Evol. Microbiol.* **50**, 1931–1942 (2000).
45. Bergthorsson, U., Adams, K. L., Thomason, B. & Palmer, J. D. Widespread horizontal transfer of mitochondrial genes in flowering plants. *Nature* **424**, 197–201 (2003).
46. Blair, J. E., Ikeo, K., Gojobori, T. & Hedges, S. B. The evolutionary position of nematodes. *BMC Evol. Biol.* **2**, 7 (2002).
47. Thompson, J. D., Higgins, D. G. & Gibson, T. J. Clustal-W—improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
48. Hall, T. A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **41**, 95–98 (1999).
49. Swofford, D. L. *PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods) (Version 4.0b10)* (Sinauer, Sunderland, Massachusetts, 2002).
50. Posada, D. & Crandall, K. A. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**, 817–818 (1998).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We are grateful to P. Cliften, M. Johnston and the Washington University Genome Sequencing Center for access to genome sequence data for *S. kudriavzevii*, *S. castellii* and *S. kluyveri*; the staff of the *Saccharomyces* Genome Database (<http://www.yeastgenome.org/>) for access to genome sequence data for *S. cerevisiae*, *S. paradoxus*, *S. mikatae* and *S. bayanus*; and the Stanford Genome Technology Center website (<http://www-sequence.stanford.edu/group/candida>) for access to sequence data for *C. albicans*. We thank D. Baum, B. Hersh, C. Hittinger and K. Johnson for useful comments on the manuscript, D. Baum and members of the Carroll laboratory for useful discussions on phylogenetics, and D. Lautenschlager for computer support. A.R. is a Human Frontier Science Program long-term fellow, B.L.W. and N.K. are NIH post-doctoral fellows, and S.B.C. is an investigator of the Howard Hughes Medical Institute. This work was funded by the Howard Hughes Medical Institute.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.B.C. (sbcarrol@wisc.edu).

The DNA sequence and analysis of human chromosome 6

A. J. Mungall*, S. A. Palmer, S. K. Sims, C. A. Edwards, J. L. Ashurst, L. Wilming, M. C. Jones, R. Horton, S. E. Hunt, C. E. Scott, J. G. R. Gilbert, M. E. Clamp, G. Bethel, S. Milne, R. Ainscough, J. P. Almeida, K. D. Ambrose, T. D. Andrews, R. I. S. Ashwell, A. K. Babbage, C. L. Bagguley, J. Bailey, R. Banerjee, D. J. Barker, K. F. Barlow, K. Bates, D. M. Beare, H. Beasley, O. Beasley, C. P. Bird, S. Blakey, S. Bray-Allen, J. Brook, A. J. Brown, J. Y. Brown, D. C. Burford, W. Burrill, J. Burton, C. Carder, N. P. Carter, J. C. Chapman, S. Y. Clark, G. Clark, C. M. Clee, S. Clegg, V. Cobley, R. E. Collier, J. E. Collins, L. K. Colman, N. R. Corby, G. J. Coville, K. M. Culley, P. Dhami, J. Davies, M. Dunn, M. E. Earthrowl, A. E. Ellington, K. A. Evans, L. Faulkner, M. D. Francis, A. Frankish, J. Frankland, L. French, P. Garner, J. Garnett, M. J. R. Ghorl, L. M. Gilby, C. J. Gillson, R. J. Glithero, D. V. Grafham, M. Grant, S. Gribble, C. Griffiths, M. Griffiths, R. Hall, K. S. Halls, S. Hammond, J. L. Harley, E. A. Hart, P. D. Heath, R. Heathcott, S. J. Holmes, P. J. Howden, K. L. Howe, G. R. Howell, E. Huckle, S. J. Humphray, M. D. Humphries, A. R. Hunt, C. M. Johnson, A. A. Joy, M. Kay, S. J. Keenan, A. M. Kimberley, A. King, G. K. Laird, C. Langford, S. Lawlor, D. A. Leongamornlert, M. Liversha, C. R. Lloyd, D. M. Lloyd, J. E. Loveland, J. Lovell, S. Martin, M. Mashreghi-Mohammadi, G. L. Maslen, L. Matthews, O. T. McCann, S. J. McLaren, K. McLay, A. McMurray, M. J. F. Moore, J. C. Mullikin, D. Niblett, T. Nickerson, K. L. Novik, K. Oliver, E. K. Overton-Larty, A. Parker, R. Patel, A. V. Pearce, A. I. Peck, B. Phillimore, S. Phillips, R. W. Plumb, K. M. Porter, Y. Ramsey, S. A. Ranby, C. M. Rice, M. T. Ross, S. M. Searle, H. K. Sehra, E. Sheridan, C. D. Skuce, S. Smith, M. Smith, L. Spraggon, S. L. Squares, C. A. Steward, N. Sycamore, G. Tamlyn-Hall, J. Tester, A. J. Theaker, D. W. Thomas, A. Thorpe, A. Tracey, A. Tromans, B. Tubby, M. Wall, J. M. Wallis, A. P. West, S. S. White, S. L. Whitehead, H. Whittaker, A. Wild, D. J. Willey, T. E. Wilmer, J. M. Wood, P. W. Wray, J. C. Wyatt, L. Young, R. M. Younger, D. R. Bentley, A. Coulson, R. Durbin, T. Hubbard, J. E. Sulston, I. Dunham, J. Rogers & S. Beck*

The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

* These authors contributed equally to this work

Chromosome 6 is a metacentric chromosome that constitutes about 6% of the human genome. The finished sequence comprises 166,880,988 base pairs, representing the largest chromosome sequenced so far. The entire sequence has been subjected to high-quality manual annotation, resulting in the evidence-supported identification of 1,557 genes and 633 pseudogenes. Here we report that at least 96% of the protein-coding genes have been identified, as assessed by multi-species comparative sequence analysis, and provide evidence for the presence of further, otherwise unsupported exons/genes. Among these are genes directly implicated in cancer, schizophrenia, autoimmunity and many other diseases. Chromosome 6 harbours the largest transfer RNA gene cluster in the genome; we show that this cluster co-localizes with a region of high transcriptional activity. Within the essential immune loci of the major histocompatibility complex, we find *HLA-B* to be the most polymorphic gene on chromosome 6 and in the human genome.

Following the announcement of the completion of the human genome project on 14 April 2003, we present here our findings on the mapping, sequencing and analysis of chromosome 6. Chromosome 6 was best known for the major histocompatibility complex (MHC), a region of 3.6 megabases (Mb) on band 6p21.3 of the short arm. The MHC has an essential role in the innate and adaptive immune system, and is characterized by high gene density, high polymorphism and high linkage disequilibrium. Much of what we know today about genetic variation and the organization of haplotypes was first discovered from studies of this region. At a time when genetic variation was assessed by serology rather than sequence, the term 'haplotype' was first introduced to describe "the combination of individual antigenic [MHC] determinants that are positively controlled by an allele"¹. Because of its crucial role in immunity and its association with many common diseases, the MHC was sequenced well ahead of the rest of chromosome 6 (ref. 2).

Particular care was taken to ensure that the highest quality was achieved for the sequence, analysis and annotation of chromosome 6. The annotation of all gene structures was manually checked and, in some cases, led to the correction of known reference genes. In addition to the genome sequences of *Mus musculus* and *Tetraodon nigroviridis*, the comparative analysis was enhanced by the inclusion (for the first time in the analysis of human chromosomes) of the recently assembled genomes of *Rattus norvegicus*, *Fugu rubripes* and *Danio rerio*. Our analysis is available through the new vertebrate genome annotation (VEGA) database (<http://vega.sanger.ac.uk/>),

making the chromosome 6 annotation a high-quality and instantly available resource.

Clone map and sequence map

Bacterial clone contigs were assembled using restriction enzyme fingerprinting and sequence-tagged site (STS) content analysis of the clones, anchored to a radiation hybrid (RH) map with a marker density of 16 per Mb. A tiling path of 1,797 clones and polymerase chain reaction (PCR) fragments (see Supplementary Table S1) were selected for sequencing spanning the chromosome in nine contigs separated by gaps of 50–200 kilobases (kb), as estimated by DNA fibre fluorescence *in situ* hybridization (FISH) (see Supplementary Table S2). All but two gaps (gaps 2 and 6) reside in the pericentromeric or sub-telomeric chromosomal regions. We assessed the chromosome coverage in several ways. First, 38% of the clones selected for sequencing were hybridized to metaphase chromosomes using FISH. This provided independent support of the map construction and also highlighted the presence of intra- and inter-chromosomal repeats. Next we identified known chromosome 6 markers in both genetic (deCODE³ and Marshfield comprehensive genetic maps⁴) and RH maps ($n = 3,036$). D6S1694 was the only genetic marker found to be absent from the sequence. The position of D6S1694 on these maps indicates that it is likely to reside within gap 6, between the sequences AL135906 and AL731777. We also accounted for all RefSeq genes mapping to chromosome 6. In the final sequence, no RefSeq gene was entirely missing. Three RefSeq

genes were found to be only partially present in the sequence (T. Furey, personal communication). Two of these, *LATS1* (NM_004690) and *C6orf35* (NM_018452), extended into sequence gaps for which there is now unfinished sequence that is confirmed to contain them: BX322632 and BX284653, respectively. Alignment of the *ASF1A* (NM_014034) messenger RNA to the genomic sequence indicates a potential deletion/polymorphism event in the P1-derived artificial chromosome (PAC) RP3-329L24 (AL132874.30), which requires further investigation. Finally, as a further check, we examined fosmid and bacterial artificial chromosome (BAC) end-sequences aligned to the chromosome 6 sequence in the University of California, Santa Cruz Genome Browser (<http://genome.cse.ucsc.edu/>) to verify the final sequence assembly.

Of the 1,797 tiling path clones, 1,739 were sequenced and finished at the Sanger Institute. For a small number of clones, the initial draft sequence was carried out by others who are credited in the corresponding EMBL (European Molecular Biology Laboratory Nucleotide Sequence Database)/GenBank/DBJ (DNA Data Bank of Japan) database submissions. The remaining 58 clone sequences have all been published previously (see Supplementary Table S3). At the telomeres, half-YAC (yeast artificial chromosome) analysis determined that the sequence extends to within 5 kb and 3 kb of the (TTAGGG)_n telomeric repeat motif for the shortest known allele of the p and q arms, respectively (H. Riethman, personal communication; <http://www.wistar.upenn.edu/Riethman/>). We detected alpha satellite repeat sequences flanking either side of the centromere, but the higher-order alpha satellite repeat structure characteristic of the core centromere has only been reached for the long arm. The finished chromosome sequence is 166,880,988 base pairs (bp) in length, and we estimate that this represents 99.5% of the euchromatin. Independent quality assessment⁵ confirmed the sequence to be at least 99.99% accurate. The current sequence assembly (v1.4) and that used in the analysis presented in this study (v1.3) are available at <http://www.sanger.ac.uk/HGP/Chr6/>.

Gene index

Crucial to the utility of a reference sequence is the associated annotation. On the basis of human interpretation of supporting evidence, we annotated 2,190 gene structures on the finished sequence (<http://vega.sanger.ac.uk/>). These gene loci were classified as previously defined⁶. Briefly, we identified 772 'known genes', identical to known human complementary DNAs or protein sequences; 287 'novel genes', with an open reading frame (ORF) and identity to spliced expressed sequence tags (ESTs); 213 'novel transcripts', as for novel genes but with an ambiguous ORF; 285 'putative genes', with identity to spliced ESTs not containing an ORF; and 633 pseudogenes, similar to known proteins but containing frameshifts and/or stop codons that disrupt the ORF. We searched for CpG islands 2 kb upstream and 1 kb downstream of the 5' and 3' ends of each annotated gene, and found that 61% of known genes were associated with a CpG island, consistent with previous estimates.

Table 1 summarizes the main features of the annotation. Excluding pseudogenes, the overall chromosome 6 gene density is 9.2 genes

per Mb; after accounting for the gene-rich MHC (43 genes per Mb), it represents a relatively gene-poor chromosome. The length of sequence occupied by annotated genes (including introns but excluding pseudogenes) is 70,396,075 bp or 42.2% (mean gene length is 31,195 bp). This is comparable to previous estimates for chromosomes 7, 14, 20 and 22 (46.5%, 43.6%, 42.4% and 51%, respectively). The chromosome sequence occupied by exons is only 2.2% with a mean exon length of 281 bp. The longest coding exon (9,114 bp) belongs to the known gene *ZNF451*; the *BPAG1* gene has the most exons with 101. The longest intron is the first intron of the *TCBA1* gene, spanning 479 kb. The largest gene is the *PARK2* gene at 6q24, which spans almost 1.4 Mb, has 12 exons and is mutated in patients with juvenile onset parkinsonism⁷. Finally, the mean number of transcripts annotated per gene is 2.34 (excluding putative genes), and the *FYN* oncogene has the most with 16 annotated transcripts.

Compared with protein-coding genes, non-protein-coding RNA genes are difficult to predict and verify experimentally. They can be divided into a growing number of classes involved principally in structural, catalytic or regulatory function⁸. tRNA genes are one of the functionally best-understood classes and can be predicted with high confidence. We determined the distribution of 616 tRNA genes across the human genome (Fig. 1). The most striking cluster contains 157 tRNAs, including all major species except for Asn-tRNA and Cys-tRNAs, and is located on chromosome 6p within the extended MHC class I region (shown as a double peak in Fig. 1). Other notable clusters are located on chromosomes 1, 5, 7, 14, 16 and 17. Except for the cluster on chromosome 7 (90% Cys-tRNAs), all these clusters contain different mixtures of tRNA genes, excluding tandem duplication of individual tRNAs as the main mechanism for their origin. RNAs including tRNAs are required in enormous quantities in the cell, constituting about 80% of cellular transcription in eukaryotes⁹. We tested whether tRNA genes co-localize with chromosomal regions that have higher-than-average transcriptional activity (transcription hotspots) of other genes, by identifying putative transcription hotspots using a database of non-normalized human ESTs. Disregarding the effects of RNA turnover, this analysis revealed the major peaks of transcription shown in Fig. 1. Sixty per cent of the tRNA clusters co-localize with predicted transcription hotspots, including the MHC, and an exact randomization test¹⁰ shows that this association is highly nonrandom ($P < 0.001$; see Methods). We postulate that selection-mediated recruitment and/or hitchhiking might be mechanism(s) responsible for the observed co-localization. This is another demonstration of the concept that chromosomal location matters, as has previously been emphasized by studies showing large-scale (multi-Mb) changes in chromatin organization following transcriptional activation¹¹.

Chromosome features

Figure 2 and Supplementary Fig. S1 summarize all the features identified on chromosome 6 in our analysis. Supplementary Fig. S1 provides a detailed view, including separate tracks for tiling path clones, genetic markers, various types of repeat, G+C content, CpG

Table 1 Gene classification and statistics

Category *	Number of genes	Total gene length (bp)	Mean gene length (bp)	Mean exon length (bp)	Mean transcripts per gene	Mean exons per gene
Known genes	772	50,364,469	65,239	303	2.78	9.95
Novel CDS	287	10,180,377	35,472	309	1.86	5.81
Novel transcripts	213	7,058,938	33,141	262	1.40	3.25
Known and novel genes	1,272	67,603,784	53,148	301	2.34	7.90
Putative genes	285	2,792,291	9,798	248	1.13	2.36
Non-pseudogenes	1,557	70,396,075	45,213	298	2.12	6.88
Pseudogenes	633	844,936	1,335	568	1.00	1.34
Total gene structures	2,190	71,241,011	32,530	318	1.79	5.28

* As defined by Deloukas *et al.*⁶ and at <http://vega.sanger.ac.uk/>.

content, single-nucleotide polymorphisms (SNPs, random and total), CpG islands, transcription start sites, similarity to mouse, rat, *Fugu*, *Tetraodon* and zebrafish, gene clusters and genes/pseudo-genes.

Repeat sequences are the most abundant features shaping the

chromosomal landscape, providing a detailed 'fossil record' of the chromosome. The repeat content of chromosome 6 is 43.95%. Transposon-derived repeats such as long interspersed elements (LINEs), short interspersed elements (SINEs), long terminal repeat (LTR) retrotransposons and DNA transposons occupy 20.85%,



Figure 1 Distribution of tRNA genes and transcription hotspots in the human genome. The red bars to the right of the ideogram for each human chromosome represent the tRNA

count per 2 Mb of sequence. The blue bars represent the number of non-normalized EST matches mapped to Ensembl genes in a 2 Mb window.

11.29%, 8.05% and 3.24% of the sequence, respectively. The repeat analysis also revealed an unusually striking structure for the human genome sequence, consisting of 17 tandemly arranged, adjoining *Alu* repeat fragments within the clone RP11-13J16 (AL499606) in 6pter, indicative of a complex series of duplications. There is mounting evidence that *Alu* repeat clusters are mediators of recurrent chromosomal aberrations¹², and it is interesting to note that there are a number of disease phenotypes involving chromosome rearrangements at 6pter, including glaucoma¹³, orofacial cleft palate¹⁴ and neoplasms (<http://cgap.nci.nih.gov/Chromosomes/Mitelman/>).

The sex-averaged genetic length of chromosome 6 in the deCODE map³ is 189.60 cM, giving a chromosome average recombination rate of 1.11 cM per Mb. In a previous study, Yu and colleagues¹⁵ identified only one recombination 'desert' on chromosome 6 (defined as a sex-averaged recombination rate of ≤ 0.3 cM per Mb for physical distances up to 5 Mb in length) between markers D6S1706 and D6S1608, and no recombination 'jungles' (>3 cM per Mb). However, analysis of the deCODE map, which offers a fivefold increase in resolution over previous genetic maps and high-resolution recombination maps in the MHC based on single-sperm typing^{16,17}, poses some questions about the desert and jungle concept. The classical MHC, a 3.6 Mb region at 6p21.3, has an overall low sex-averaged recombination rate across its length in the genetic map (0.49 cM per Mb; Fig. 3) and yet three hotspots of recombination are observed at high resolution¹⁷ (Fig. 3, inset). The finding of recombination hotspots within a region of low long-range recombination rate illustrates that even the current genome-wide genetic map masks local recombination hotspots/rates, and hence the description of Mb regions as containing high or low recombination may be of limited value in understanding recombination rates. Despite this, we were able to identify several small regions of interest. We found that 116 marker intervals had recombination rates greater than the chromosome average (1.11 cM per Mb), ten of these showing greater than a fivefold increase and four intervals showing a tenfold increase in recombination rate (Table 2). In the four intervals with the highest recombination rates, the markers were within 20 kb of each other in the sequence. It could be that these intervals pin-point the location of recombination events and provide indicators of possible hotspot locations, although genotyping errors could also result in inflated recombination rates within these intervals.

Recent studies suggest that around 5% of the human genome has arisen from segmental duplications¹⁸. We initially identified large segmental duplications on chromosome 6 by the observed 'stacking' of restriction enzyme fingerprints in the clone assembly as a result of near-identical sequences (and therefore restriction sites). One such duplication lies on 6p, which harbours two pseudogenes of the ancestral β -glucuronidase (*GUSB*) gene on 7q11.21, one in the extended MHC at 6p21.31, and the second in the pericentromeric region 6p11.1. The spinal muscular atrophy (*SMA*) locus of chromosome 5q13 and three other chromosome 5 regions also

contain paralogous *GUSB* sequences. Clones mapping to these regions by fingerprint and STS-content analysis (for example, RP11-239L20 and b55C20) were used in FISH to metaphase chromosomes and showed multiple signals on chromosomes 6, 5 and 22 (Supplementary Table S4), in agreement with recently published data¹⁹.

Both large- and small-scale duplications are required to explain the age distribution of human gene families²⁰. To assess the effect of local duplications on the number of genes on chromosome 6, we carried out a gene cluster analysis (excluding pseudogenes) using BLASTP. For the purpose of this analysis, a cluster was defined as two or more paralogous genes (e -values $<1 \times 10^{-15}$) within 1 Mb of each other. Adjacent clusters of the same gene family were further grouped into superclusters. In this analysis, we found 65 clusters (Supplementary Fig. S1 and Supplementary Table S5), of which 25 could be grouped into six superclusters, resulting in 46 gene family clusters containing 2–55 paralogues per cluster. In all, 223 (14.3%) out of the 1,557 annotated genes on chromosome 6 seem to have arisen through local duplication. This number is likely to increase if global (genome-wide) duplication events are considered as well. Interestingly, all major gene clusters locate to the extended MHC region on the short arm of chromosome 6 (ref. 21), with the exception of the recently described *RAET* genes²². The remaining 39 clusters contain two to four paralogues each and are scattered randomly across chromosome 6. In addition, we carried out a cluster analysis based on known protein domains using InterProScan (<http://www.ebi.ac.uk/interpro/scan.html>), resulting in a similar, although not identical, ranking of top-scoring clusters (Supplementary Table S6).

Comparative analysis

The recent accumulation of assembled genome sequences for mouse, rat, *Tetraodon*, *Fugu* and zebrafish has allowed us to evaluate our gene annotation. The results are summarized in Table 3 and Supplementary Table S7. Because comparative genome analysis was not used in the gene structure annotation, it is possible to use it to estimate the completeness of annotation by considering regions that are conserved in all six organisms (the assumption being that such completely conserved regions are unlikely to be the result of artefact). There are 5,409 such regions, of which 5,204 have overlap with 4,466 annotated exons, suggesting that 95.6% (4,466/(4,466 + 205)) of coding exons on chromosome 6 have been annotated. This estimate is likely to be low because not all sequence conservation will be confined to exons. The figure is, however, in agreement with similar estimates of annotation coverage made for chromosomes 20 (ref. 6) and 14 (ref. 23), and to a lesser extent chromosome 22 (ref. 24), although slightly different methods and resources were used in each case.

It is also interesting to consider the conservation between chromosome 6 and groups of species. For example, regions of the human sequence which are conserved in either rat, mouse, *Fugu*, zebrafish or *Tetraodon* account for more exons than regions of conservation with any single species in isolation: 84% of genes and 80% of exons (98% and 88% respectively if we consider only "Known genes") are supported by conservation with at least two

Table 2 Inter-marker recombination rates

Marker left	Marker right	Interval size (bp)	Sex-averaged position in deCODE map (cM)	Recombination rate (cM per Mb)
D6S1038	D6S1722	6,245	132.25	59.25
D6S383	D6S970	6,284	145.75	33.42
D6S444	D6S417	3,677	99.71	19.04
D6S1571	D6S1545	19,075	49.26	12.58
D6S1588	D6S1029	32,586	44.86	10.74
D6S1578	D6S1559	132,974	33.27	10.08
D6S407	D6S1690	20,782	127.89	9.62
D6S464	D6S105	39,531	50.57	8.09
D6S942	D6S1600	65,330	0	6.89
D6S1697	D6S503	193,688	184.1	5.89

Figure 2 Features of the human chromosome 6 sequence. Feature tracks from left to right are as follows. (1) Sequence scale in megabases. (2) Extent of human chromosome 6 sequence (black) separated by gaps (grey). (3) Rodent synteny: mouse, left; rat, right. (4) Location of predicted CpG islands. (5) From left to right, the location of sequence homology to *Fugu*, zebrafish and *Tetraodon*. (6) The location of the 'known' and 'novel CDS' (novel coding sequence) annotated gene structures. Official gene symbols are used when available. Because of space limitations, the foldout shown in this figure had to be condensed for the printed version. Therefore, we strongly recommend the downloading of Supplementary Fig. S1 to follow points discussed in the main text.

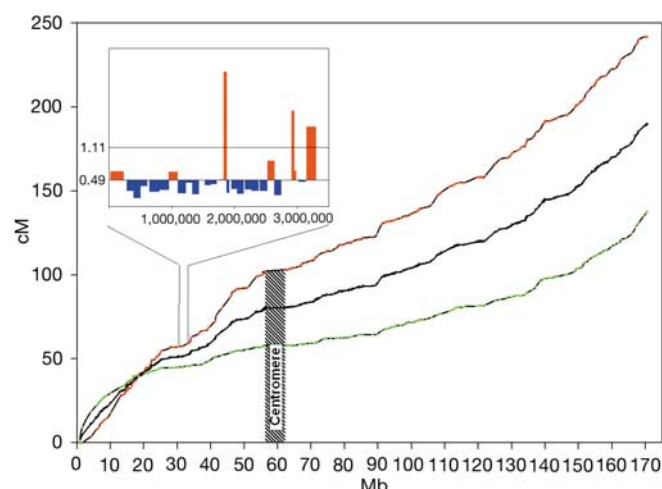


Figure 3 Alignment of the genetic markers in the deCODE linkage map³ with the chromosome 6 sequence. The physical position of each genetic marker on the female, the male and the sex-averaged genetic map is indicated. Inset, high-resolution recombination intensity across the 3.6 Mb classical MHC region at 6p21.3 (ref. 17).

of the above species, 81% of genes and 77% of exons when considering the best single species comparison (mouse) (see Supplementary Information). Figure 4 illustrates this point and emphasizes the potential of multi-species comparative analysis²⁵ for the identification of evolutionarily conserved regions (ECRs). In this case, the three identified ECRs have an open reading frame and putative splice sites suggesting that they constitute part of a potential novel gene.

Sequence variation

The most common variations in the human genome are SNPs, and their exploration and use is expected to revolutionize our understanding of human disease and evolution²⁶. Towards this end, we have mapped a total of 183,019 SNPs onto the finished sequence of chromosome 6. The non-normalized SNPs were taken from the dbSNP database (<http://www.ncbi.nlm.nih.gov/SNP/>), where they have been deposited by the International SNP Consortium (TSC)²⁷ and others. Of these SNPs, 113,284 (62%) were discovered as part of the chromosome 6 project itself by sequence comparison of clone overlaps using a modification of the SSAHA program²⁸. After correlating the positions of all SNPs to the annotation presented here, chromosome 6 contains 2,761 SNPs in protein-coding exons. When applied genome-wide (using the Ensembl annotation), followed by further clustering of the SNPs according to gene structures and normalizing for coding-sequence length, this analysis provides an estimate of the most polymorphic genes in the genome. With 86 SNPs per kb mapping to *HLA-B* (human leukocyte antigen, class I, B), we determined this MHC class I gene to be the most polymorphic gene on chromosome 6 and in the human genome, closely followed by other class I and class II genes. This result is concordant with the HLA database (<http://www.ebi.ac.uk/imgt/hla/stats.html>), which lists *HLA-B* as the most polymorphic MHC gene, with 511 known alleles.

Supplementary Fig. S1 shows the distribution of SNPs across chromosome 6. The lower track shows the total SNP density but is likely to be skewed by local SNP-discovery efforts. The upper SNP track shows the distribution of a random set of SNPs generated by the TSC and is, therefore, a better indicator of regions of high and low sequence variation. As expected from previous studies², the most variable regions map to segments containing MHC class I and class II genes. Other notable regions of high sequence variation are located at 26.7–27.0 Mb, 57.4–57.5 Mb and 58.2–58.5 Mb on the short arm of the chromosome. Interestingly, all three regions contain segmental duplications, supporting the hypothesis that paralogous sequence variants may have been falsely identified as SNPs in dbSNP¹⁸, thereby increasing the apparent density of ‘SNPs’.

Medical implications

At the time of writing, there are 130 genes mapping to chromosome 6 that cause, predispose to or protect from disease (Supplementary Table S8). Of these, 84 (65%) have already been cloned and include the well-studied MHC class I-like gene, *HFE*²⁹, mutated in hereditary haemochromatosis. With respect to autoimmunity, the MHC is the most important genetic region in the human genome. Well over 100 diseases have been linked to the MHC, including most if not all autoimmune diseases³⁰. Autoimmune disorders are common, complex, and involve genetic and environmental factors³¹. They affect about 4% of the population, and include type 1 diabetes, rheumatoid arthritis and multiple sclerosis, to name but a few.

Genes with protein products that affect the brain or neural tissue are of particular interest in the study of neurological or psychiatric disorders. Examples of chromosome 6 genes that have been cloned include *SCA1*, which can cause spinocerebellar ataxia when mutated; *EPM2A*, a gene mutated in patients with Lafora’s myoclonus epilepsy³²; and the *PARK2* gene, point mutations or deletions in which are responsible for a juvenile-onset autosomal-recessive form of parkinsonism⁷. The application of genetic mapping to human disease gene identification has led to the definition of many linkages for, or associations with, complex traits on chromosome 6. One such trait is schizophrenia. Recent studies^{33,34} found family-based association with the dysbindin (*DTNBP1*) gene at 6p22. Dysbindin is a promising candidate because it is localized to presynaptic terminals, and could be involved in the formation and/or maintenance of synapses and in signal transduction. *HTR1E*, at 6q14.2, and *B3GAT1*, at 11q25, have recently been identified as candidate genes for a schizophrenia-like psychosis through molecular analysis of a balanced translocation t(6:11)(q14.2;q25)³⁵. Sequence analysis of the translocation breakpoints reveals that neither gene is disrupted; however, the expression of these genes could potentially be altered by a positional effect due to alteration of the chromatin environment as a consequence of the translocation.

The complete sequence of chromosome 6 not only allows us to look at genetic conditions, but also provides a tool with which to study epigenetic changes resulting in disease³⁶. One such disease is transient neonatal diabetes mellitus (TNDM), which is the result of overexpression of an imprinted and paternally expressed gene(s) at 6q24. Two imprinted genes at 6q24, *PLAGL1* and *HYMAI*, have been identified as potential candidates for TNDM through studies of paternal uniparental isodisomy of chromosome 6 (UPD6), paternally inherited duplication of 6q24, and a methylation defect at a

Table 3 Comparative sequence analysis of the annotation set

Species	Known		Novel CDS		Novel transcript		Pseudogene		Putative		Total	
	G	E	G	E	G	E	G	E	G	E	G	E
Rodents and fish	545 (0.71)	3,495 (0.33)	163 (0.57)	567 (0.26)	15 (0.07)	27 (0.03)	319 (0.50)	345 (0.41)	9 (0.03)	32 (0.04)	1,051 (0.48)	4,466 (0.30)
Rodents or fish	757 (0.98)	9,214 (0.88)	275 (0.96)	1,776 (0.82)	108 (0.51)	226 (0.26)	592 (0.94)	719 (0.85)	111 (0.39)	175 (0.24)	1,843 (0.84)	12,110 (0.80)

The number of annotated genes and exons in each category supported by an overlapping conserved region. A more detailed breakdown of the comparative sequence analysis by species is given in Supplementary Information. G, Genes; E, Exons.

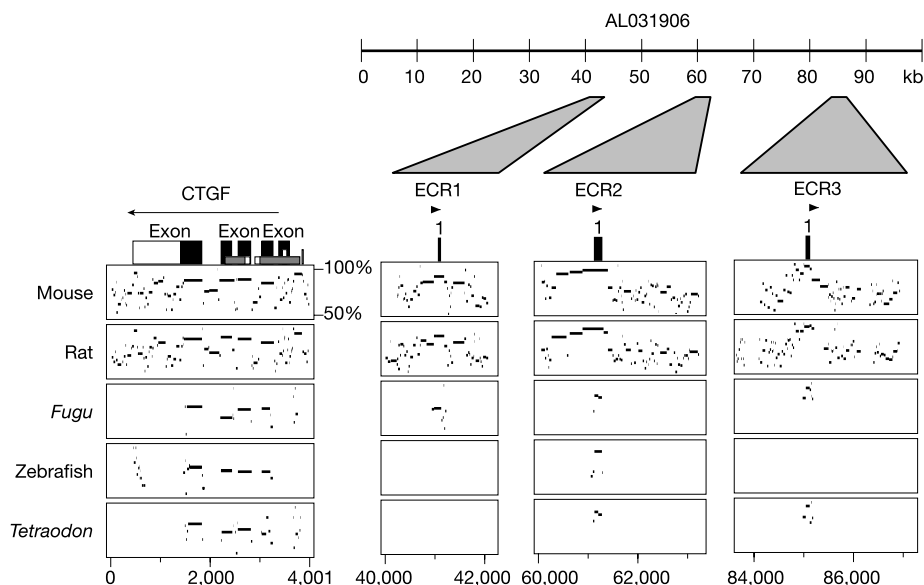


Figure 4 Identification of evolutionarily conserved regions using multi-species comparative analysis. Pair-wise alignments between the chromosome 6 sequence and the draft sequences of each indicated species were generated and displayed using MultiPipMaker⁴⁹. The percentage identity of each alignment is indicated. The first plot

shows the known connective tissue growth factor (CTGF) gene aligned to syntenic draft sequences of rodents and fish. Three ECRs identified from the comparative sequence analysis were found within 50 kb of sequence AL031906.

CpG island. A second cluster of imprinted genes occurs at 6q26, a region actively studied for the presence of tumour-suppressor genes (TSGs).

Tumour-specific mutations that inactivate TSGs often involve large-scale changes, such as loss of the entire chromosome or large fragments of it containing the TSG. The long arm of chromosome 6 has been a focus of attention for cancer geneticists because it harbours genes of importance in tumour progression in a diverse set of solid and haematological malignancies (see ref. 37 and references therein). DNA, amplified by PCR, from the clones used in the sequencing of chromosome 6 have been arrayed for comparative genomic hybridization, and are currently being used to define commonly deleted and/or homozygous deletions in astrocytic gliomas, the results of which will be published elsewhere. The use of sequenced clones in the array provides a direct link back to the sequence and its features, and therefore provides a new and powerful tool in the search for TSGs. Furthermore, the identification of the genetic loci and associated polymorphisms responsible for complex traits should greatly benefit from the availability of the complete sequence. □

Methods

The strategy and method of construction of the bacterial clone physical map of chromosome 6 has been described elsewhere^{38,39}. Briefly, 2,802 STSs on the radiation hybrid map of chromosome 6 (<http://www.sanger.ac.uk/cgi-bin/rhmap?chr=6>) were used to screen up to 87 genomic equivalents (87× coverage) of BAC, PAC, cosmid and YAC libraries. Identified clones were fluorescently fingerprinted and their landmark content established to generate clone contigs. End-sequencing of terminal clones in the contigs was performed and novel STSs were designed for further walking experiments. In regions devoid of bacterial clone coverage, STSs flanking the gaps were used to screen YAC libraries. Multiple checks were performed on each clone selected for sequencing from the map.

Random shotgun sequences of bacterial clones were generated from pUC plasmids with inserts of mainly 1.4–4 kb, which were sequenced from both ends using the dideoxy chain termination method with different versions of big dye terminator chemistry⁴⁰. The resulting sequencing reactions were analysed on various models of ABI sequencing machines, and the data generated were processed by a suite of in-house programs (<http://www.sanger.ac.uk/Software/sequencing/>) before assembly with the PHRED and PHRAP (<http://www.phrap.org/>) algorithms. For the finishing phase, we used the GAP4 program⁴¹ to help us to assess, edit and select reactions, to eliminate ambiguities and to close sequence gaps. All DNA sequence has been deposited in the EMBL, GenBank or DDBJ databases.

The finished genomic sequence was analysed using an automatic Ensembl pipeline⁴² with modifications to aid the manual curation process. The identification of interspersed repeats using RepeatMasker; simple repeats using Tandem Repeat Finder⁴³; matches to vertebrate cDNAs and ESTs using WU-BLASTN and EST_GENOME; and *ab initio* gene prediction using FGENESH and GENESCAN was as described previously⁶. A protein database combining non-redundant data from SwissProt and TrEMBL was also searched using WU-BLASTX. Using these pipeline results as a starting point, gene structures were manually annotated according to the human annotation workshop (HAWK) guidelines (<http://www.sanger.ac.uk/HGP/havana/hawk.shtml>) and, where possible, were given gene symbols approved by the HUGO and HLA Gene Nomenclature Committees^{44,45}. The sequence was further analysed to identify putative tRNA genes using tRNAscan-SE⁴⁶.

For comparison with mouse and rat, the repeat-masked sequence of chromosome 6 was aligned to these genomes using BLASTZ⁴⁷. The resulting matches were post-processed using the programs *axtBest* and *subsetAxt*, as previously described⁴⁷, to select the best match and to make the alignments relatively specific to exons by using a specific scoring matrix and threshold. For the comparative analysis with fish, genome sequences were aligned to the chromosome using WU-TBLASTX, with the scoring matrix, parameters and filtering strategy used in the Exofish (exon finding by sequence homology) procedure⁴⁸. Overlapping alignments to different sequences were merged to produce contiguous regions of sequence conservation, analogous to the evolutionarily conserved regions, or 'Ecores', reported by Exofish.

SNPs determined as part of this sequencing project were identified from sequence overlaps using a modification of the SSAHA program²⁸. Clone overlaps from available finished and unfinished public human genomic sequence were aligned, and high-quality base discrepancies ($Q \geq 23$) were identified as candidate SNPs provided that the total overlap was >6,000 bp. Candidate SNPs were rejected if any of its neighbouring five bases had Phrap quality values of <15 (bases of finished sequence were assumed to have a quality value of 40; that is, one error in 10,000 bp) or if fewer than nine of the ten neighbours matched. If the number of detected SNPs in one clique was greater than five in a 500 bp interval, then all SNPs were discarded for that interval. All SNPs for a clone overlap were similarly discarded if the SNP density for the entire overlap region was less than one SNP per 4,000 bp. In some overlap regions, more than two clones overlap. A SNP detected from multiple overlapping clones was merged and represented as one SNP.

For transcription hotspot analysis, human ESTs derived from non-normalized libraries were matched to the human genome sequence with BLAT using the Ensembl pipeline. Matches were accepted only if they showed >97% nucleotide identity over >90% of the EST sequence; if they were the single best identity match for an individual EST; and if they exhibited more than 80% coverage with an Ensembl gene. The null hypothesis that the association of the tRNA and expression peaks is random was then tested: with the genome divided into 2 Mb windows, peaks of expression and tRNA clusters were defined as those windows where the number of matches was greater than the mean value by more than one standard deviation. The positions of the resultant tRNA peaks were randomized, and a distribution of the associations with an expression peak (either coincident or within 2 Mb of the expression peak) was generated from 10,000 replicates. The observed association of peaks was compared to this distribution with a *P*-value calculated as the proportion of replicates where an equal or greater number of tRNA peaks was associated with an expression peak.

Received 8 June; accepted 11 September 2003; doi:10.1038/nature02055.

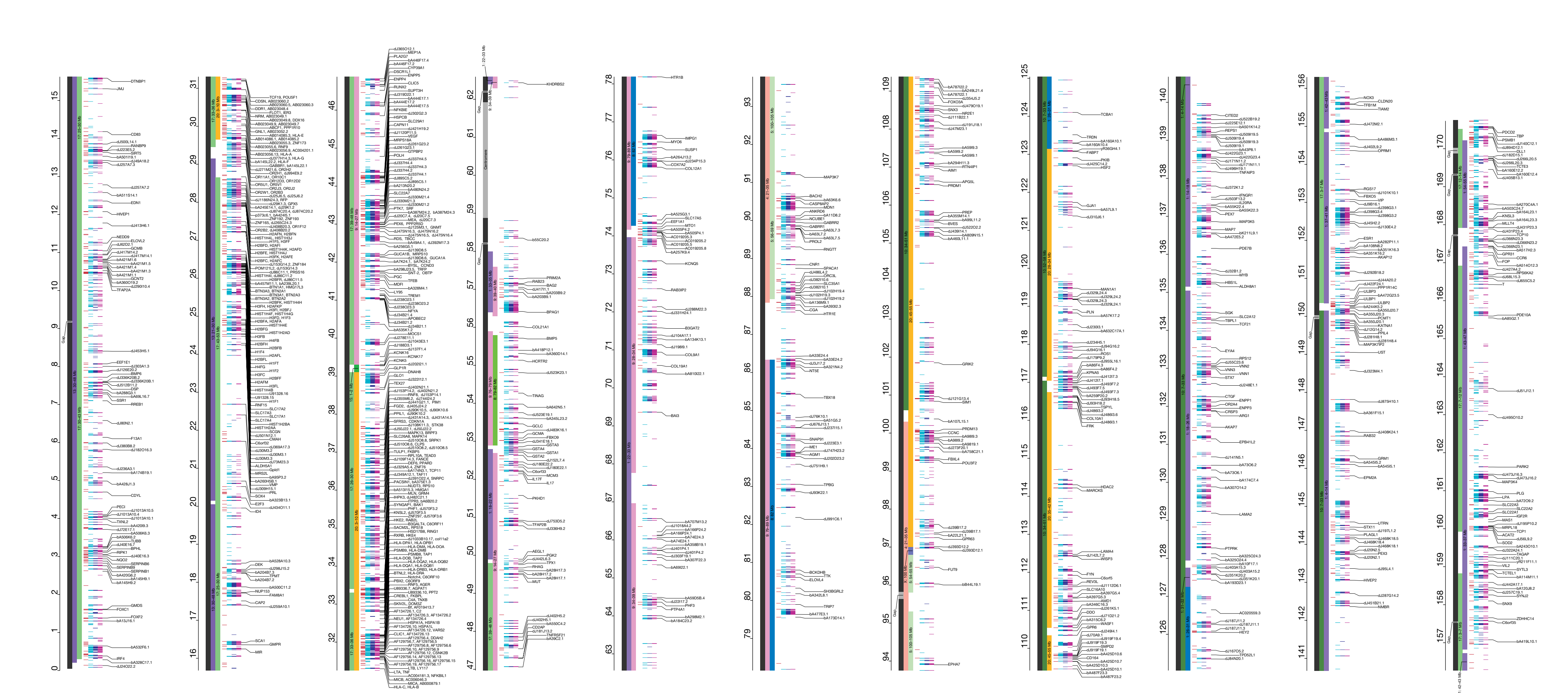
1. Cepellini, R. *et al.* in *Histocompatibility Testing* (eds Curtoni, E. S., Mattiuz, P. L. & Tosi, R. M.) 149–184 (Munksgaard, Copenhagen, 1967).
2. The MHC sequencing consortium. Complete sequence and gene map of a human major histocompatibility complex. *Nature* **401**, 921–923 (1999).
3. Kong, A. *et al.* A high-resolution recombination map of the human genome. *Nature Genet.* **31**, 241–247 (2002).
4. Broman, K. W., Murray, J. C., Sheffield, V. C., White, R. L. & Weber, J. L. Comprehensive human genetic maps: individual and sex-specific variation in recombination. *Am. J. Hum. Genet.* **63**, 861–869 (1998).
5. Felsenfeld, A., Peterson, J., Schloss, J. & Guyer, M. Assessing the quality of the DNA sequence from the Human Genome Project. *Genome Res.* **9**, 1–4 (1999).
6. Deloukas, P. *et al.* The DNA sequence and comparative analysis of human chromosome 20. *Nature* **414**, 865–871 (2001).
7. Kitada, T. *et al.* Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. *Nature* **392**, 605–608 (1998).
8. Eddy, S. R. Non-coding RNA genes and the modern RNA world. *Nature Rev. Genet.* **2**, 919–929 (2001).
9. Paule, M. R. & White, R. J. Survey and summary: transcription by RNA polymerases I and III. *Nucleic Acids Res.* **28**, 1283–1298 (2000).
10. Sokal, R. R. & Rohlf, F. J. in *Biometry* 3rd edn 803–819 (Freeman and Company, New York, 1995).
11. Volpi, E. V. *et al.* Large-scale chromatin organization of the major histocompatibility complex and other regions of human chromosome 6 and its response to interferon in interphase nuclei. *J. Cell Sci.* **113**, 1565–1576 (2000).
12. Kolomietz, E., Meyn, M. S., Pandita, A. & Squire, J. A. The role of Alu repeat clusters as mediators of recurrent chromosomal aberrations in tumors. *Genes Chromosomes Cancer* **35**, 97–112 (2002).
13. Lehmann, O. J. *et al.* Ocular developmental abnormalities and glaucoma associated with interstitial 6p25 duplications and deletions. *Invest. Ophthalmol. Vis. Sci.* **43**, 1843–1849 (2002).
14. Davies, A. F. *et al.* Evidence of a locus for orofacial clefting on human chromosome 6p24 and STS content map of the region. *Hum. Mol. Genet.* **4**, 121–128 (1995).
15. Yu, A. *et al.* Comparison of human genetic and sequence-based physical maps. *Nature* **409**, 951–953 (2001).
16. Jeffreys, A. J., Kauppi, L. & Neumann, R. Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nature Genet.* **29**, 217–222 (2001).
17. Cullen, M., Peretto, S. P., Klitz, W., Nelson, G. & Carrington, M. High-resolution patterns of meiotic recombination across the human major histocompatibility complex. *Am. J. Hum. Genet.* **71**, 759–776 (2002).
18. Bailey, J. A. *et al.* Recent segmental duplications in the human genome. *Science* **297**, 1003–1007 (2002).
19. Courseaux, A. *et al.* Segmental duplications in euchromatic regions of human chromosome 5: a source of evolutionary instability and transcriptional innovation. *Genome Res.* **13**, 369–381 (2003).
20. Gu, X., Wang, Y. & Gu, J. Age distribution of human gene families shows significant roles of both large- and small-scale duplications in vertebrate evolution. *Nature Genet.* **31**, 205–209 (2002).
21. Trowsdale, J. The gentle art of gene arrangement: the meaning of gene clusters. *Genome Biol.* **3**, COMMENT2002 (2002).
22. Radosavljevic, M. *et al.* A cluster of ten novel MHC class I related genes on human chromosome 6q24.2–q25.3. *Genomics* **79**, 114–123 (2002).
23. Heilig, R. *et al.* The DNA sequence and analysis of human chromosome 14. *Nature* **421**, 601–607 (2003).
24. Collins, J. E. *et al.* Reevaluating human gene annotation: a second-generation analysis of chromosome 22. *Genome Res.* **13**, 27–36 (2003).
25. Thomas, J. W. *et al.* Comparative analyses of multi-species sequences from targeted genomic regions. *Nature* **424**, 788–793 (2003).
26. Syvanen, A. C. Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nature Rev. Genet.* **2**, 930–942 (2001).
27. Sachidanandam, R. *et al.* A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409**, 928–933 (2001).
28. Ning, Z., Cox, A. J. & Mullikin, J. C. SSAHA: a fast search method for large DNA databases. *Genome Res.* **11**, 1725–1729 (2001).
29. Feder, J. N. *et al.* A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis. *Nature Genet.* **13**, 399–408 (1996).
30. Tiwari, J. L. & Terasaki, P. I. *HLA and Disease Associations* (Springer Verlag, Berlin, 1985).
31. Vyse, T. J. & Todd, J. A. Genetic analysis of autoimmune disease. *Cell* **85**, 311–318 (1996).
32. Minassian, B. A. *et al.* Mutations in a gene encoding a novel protein tyrosine phosphatase cause progressive myoclonus epilepsy. *Nature Genet.* **20**, 171–174 (1998).
33. Straub, R. E. *et al.* Genetic variation in the 6p22.3 gene *DTNBPI*, the human ortholog of the mouse dysbindin gene, is associated with schizophrenia. *Am. J. Hum. Genet.* **71**, 337–348 (2002).
34. Schwab, S. G. *et al.* Support for association of schizophrenia with genetic variation in the 6p22.3 gene, dysbindin, in sib-pair families with linkage and in an additional sample of triad families. *Am. J. Hum. Genet.* **72**, 185–190 (2003).
35. Jeffries, A. R. *et al.* B-1,3-Glucuronyltransferase-1 gene implicated as a candidate for a schizophrenia-like psychosis through molecular analysis of a balanced translocation. *Mol. Psychiatry* **8**, 654–663 (2003).
36. Novik, K. L. *et al.* Epigenomics: genome-wide study of methylation phenomena. *Curr. Issues Mol. Biol.* **4**, 111–128 (2002).
37. Acquati, F. *et al.* Cloning and characterization of a senescence inducing and class II tumor suppressor gene in ovarian carcinoma at chromosome region 6q27. *Oncogene* **20**, 980–988 (2001).
38. Mungall, A. J. *et al.* From long range mapping to sequence-ready contigs on human chromosome 6. *DNA Seq.* **8**, 151–154 (1997).
39. Bentley, D. R. *et al.* The physical maps for sequencing human chromosomes 1, 6, 9, 10, 13, 20 and X. *Nature* **409**, 942–943 (2001).
40. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
41. Bonfield, J. K., Smith, K. & Staden, R. A new DNA sequence assembly program. *Nucleic Acids Res.* **23**, 4992–4999 (1995).
42. Hubbard, T. *et al.* The Ensembl genome database project. *Nucleic Acids Res.* **30**, 38–41 (2002).
43. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).
44. Wain, H. M., Lush, M., Ducluzeau, F. & Povey, S. Genew: the human gene nomenclature database. *Nucleic Acids Res.* **30**, 169–171 (2002).
45. Robinson, J. *et al.* IMGT/HLA and IMGT/MHC: sequence databases for the study of the major histocompatibility complex. *Nucleic Acids Res.* **31**, 311–314 (2003).
46. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**, 955–964 (1997).
47. Schwartz, S. *et al.* Human–mouse alignments with BLASTZ. *Genome Res.* **13**, 103–107 (2003).
48. Roest Crollius, H. *et al.* Estimate of human gene number provided by genome-wide analysis using *Tetraodon nigroviridis* DNA sequence. *Nature Genet.* **25**, 235–238 (2000).
49. Schwartz, S. *et al.* PipMaker—a web server for aligning two genomic DNA sequences. *Genome Res.* **10**, 577–586 (2000).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank H. Riethman and R. Sudbrak for the provision of clones, H. M. Wain, R. C. Lovering, E. A. Bruford, M. J. Lush, M. W. Wright, V. K. Khodiyar, C. C. Talbot and S. Povey from the HUGO Gene Nomenclature Committee for official gene nomenclature, S. G. Marsh and the WHO HLA Nomenclature Committee for HLA gene nomenclature, Y. Chen for assistance with the SNP mapping, the EMBL and Ensembl database teams at the European Bioinformatics Institute, the HUGO chromosome editors R. D. Campbell, H. M. Cann, E. Jazwinska, J. Ragoussis and A. Ziegler for support throughout the project, P. Deloukas for critical reading of the manuscript, and the Wellcome Trust for financial support.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.J.M. (ajm@sanger.ac.uk) or S.B. (beck@sanger.ac.uk). Accession numbers for the sequence analysed for this paper can be found in Supplementary Fig. S1. All reported DNA sequences have been deposited in EMBL, GenBank or DDBJ.



The formation of the first low-mass stars from gas with low carbon and oxygen abundances

Volker Bromm & Abraham Loeb

Astronomy Department, Harvard University, 60 Garden Street, Cambridge, Massachusetts 02138, USA

The first stars in the Universe are predicted to have been much more massive than the Sun^{1–3}. Gravitational condensation, accompanied by cooling of the primordial gas via molecular hydrogen, yields a minimum fragmentation scale of a few hundred solar masses. Numerical simulations indicate that once a gas clump acquires this mass it undergoes a slow, quasi-hydrostatic contraction without further fragmentation^{1,2}; lower-mass stars cannot form. Here we show that as soon as the primordial gas—left over from the Big Bang—is enriched by elements ejected from supernovae to a carbon or oxygen abundance as small as ~ 0.01 – 0.1 per cent of that found in the Sun, cooling by singly ionized carbon or neutral oxygen can lead to the formation of low-mass stars by allowing cloud fragmentation to smaller clumps. This mechanism naturally accommodates the recent discovery⁴ of solar-mass stars with unusually low iron abundances ($10^{-5.3}$ solar) but with relatively high ($10^{-1.3}$ solar) carbon abundance. The critical abundances that we derive can be used to identify those metal-poor stars in our Galaxy with elemental patterns imprinted by the first supernovae. We also find that the minimum stellar mass at early epochs is partially regulated by the temperature of the cosmic microwave background.

The microphysics of cooling due to the presence of molecular hydrogen (H_2) within the first gas clouds to condense in the Universe introduces a characteristic temperature ($T_c \approx 100$ – 200 K) and hydrogen density ($n_c \approx 10^4 \text{ cm}^{-3}$) to the formation process of the first stars^{1,2}. The temperature floor is set by the energy separation of the lowest-lying rotational levels of H_2 , and the characteristic density leads to the thermalization of these levels, at which point cooling becomes less efficient. The minimum mass scale for fragmentation is approximately given by the Jeans mass, $M_J \approx 400 M_\odot (T/200 \text{ K})^{3/2} (n/10^4 \text{ cm}^{-3})^{-1/2}$ where subscript $\odot$ refers to the Sun. The final mass of a star is determined by the accretion process onto the nascent protostellar core, resulting in $M_* \approx \alpha M_J$ with $\alpha \lesssim 0.5$ (refs 5, 6). In contrast to the formation mode of massive stars (population III) at high redshifts, fragmentation is observed to favour stars below a solar mass (population I and II) in the present-day Universe. The transition between these fundamental modes is expected to be mainly driven by the progressive enrichment of the cosmic gas with heavy elements (or ‘metals’), which enable the gas to cool to lower temperatures.

The concept of a ‘critical metallicity’, Z_{crit} , has been used^{7–11} to characterize the transition between population III and population II formation modes (where Z denotes the mass fraction contributed by all heavy elements). Previous studies^{7,8,11} have only constrained this important parameter to within a few orders of magnitude, $Z_{\text{crit}} \approx 10^{-6}$ – $10^{-3} Z_\odot$, under the implicit assumption of solar relative abundances of metals. This assumption is likely to be violated by the metal yields of the first supernovae at high redshifts, for which strong deviations from solar abundance ratios are predicted^{12–17}. The cooling rate of the metals depends on their ionization state, which in turn is controlled by the ionizing backgrounds (ultraviolet and X-ray photons or cosmic rays) that were not fully considered in earlier studies.

Here we show that the transition between the above star for-

mation modes is driven primarily by fine-structure line cooling of singly ionized carbon or neutral atomic oxygen. We refine earlier estimates of Z_{crit} which did not explicitly distinguish between different coolants, by introducing separate critical abundances for carbon and oxygen, $[C/H]_{\text{crit}}$ and $[O/H]_{\text{crit}}$, respectively, where $[A/H] = \log_{10}(N_A/N_H) - \log_{10}(N_A/N_H)_\odot$. Why are C and O identified as the key species responsible for the global shift in the star formation mode? We find that under the temperature and density conditions that characterize population III star formation, the most important coolants are O I and C II, whose fine-structure lines dominate all other metal transitions¹⁸. Cooling due to molecules becomes important only at lower temperatures, and cooling due to dust grains only at higher densities^{7,11,18}.

The conditions that terminate the predominance of massive, population III stars have dramatic implications for the reionization history of the Universe. Massive stars are much more efficient at producing ionizing radiation than low mass stars, owing to their high surface temperature ($\sim 10^5$ K). In particular, metal-free stars with a mass $\gtrsim 100 M_\odot$ produce $\sim 5 \times 10^4$ hydrogen ionizing photons per baryon incorporated into them, a yield higher by an order of magnitude than obtained for a present-day mass function of stars^{19,20}. Massive stars can therefore reionize the Universe early^{21–23} at $z \approx 17$, as required by the recent detection²⁴ of large-scale polarization anisotropies in the cosmic microwave background. A strong ultraviolet flux just below the Lyman limit ($h\nu < 13.6$ eV) is predicted²⁵ to be emitted by the same stars that are responsible for the reionization of the intergalactic medium. This soft ultraviolet radiation can penetrate the neutral intergalactic medium and ionize any trace amount of neutral carbon due to its low first-ionization potential of 11.26 eV. Because carbon is highly underabundant, the ultraviolet background can ionize carbon throughout the Universe well before hydrogen reionization. Oxygen, on the other hand, will be predominantly neutral prior to reionization because its ionization potential is 13.62 eV. Cooling is mediated by excitations due to collisions of the respective metal with free electrons or hydrogen atoms. At the low fractional abundances of electrons, $x_e \lesssim 10^{-4}$, expected in the neutral (rapidly recombining) gas at $z \approx 15$, collisions with hydrogen atoms dominate. Even if the gas had initially been ionized to a much higher degree owing to the impact of the supernova ejecta, it would have recombined to the level stated above within its free-fall time: $t_{\text{ff}} \approx 5 \times 10^5 (n/10^4 \text{ cm}^{-3})^{-1/2}$ yr. This renders our analysis independent of the very uncertain hydrogen-ionizing backgrounds at high redshifts (cosmic rays or soft X-rays).

We now derive the critical C and O abundances. Our starting point is the characteristic state reached in primordial gas that cools only through molecular hydrogen. For the gas to fragment further (instead of following a slow contraction to a single massive star^{1,2}), additional cooling due to C II or O I is required. Fragmentation requires that the radiative cooling rate be higher than the free-fall compressional heating rate. We find the critical metal abundances by equating the two rates: $\Lambda_{\text{CII,OI}}(n, T) \approx 1.5 n k_B T / t_{\text{ff}}$, where k_B is Boltzmann’s constant. In evaluating this threshold condition, we solve the respective rate equations for C II (a two-level system) and O I (a three-level system), including all possible radiative and collisional transitions¹⁸. We assume that the emission of the fine-structure lines proceeds under optically thin conditions, as appropriate for the low atomic abundances considered here¹⁸. We find the critical C and O abundances by setting $n \approx n_c$ and $T \approx T_c$, and estimate the uncertainties by considering a plausible range near these values ($100 \text{ K} \lesssim T \lesssim 200 \text{ K}$).

In Fig. 1, we show the threshold C and O abundances that enable continued fragmentation for various temperature and density values. It is convenient to plot these quantities as functions of the Jeans mass ($\propto T^{3/2} n^{-1/2}$). We find critical abundance values of $[C/H]_{\text{crit}} \approx -3.5 \pm 0.1$ and $[O/H]_{\text{crit}} \approx -3.05 \pm 0.2$. Strictly speaking, these threshold levels are the required abundances in

the gas phase. If a fraction of these elements were depleted onto dust grains, then the total C and O elemental abundances would have to be correspondingly higher. Considering the large current uncertainties in modelling the properties and formation channels of dust at low metal abundances¹¹, we assume the limiting case of zero C and O depletion onto dust when comparing our theoretical predictions with observed stellar abundances.

We stress that even if sufficient C or O atoms are present to cool the gas further, there will be a minimum attainable temperature^{6,26} that is set by the interaction of the atoms with the thermal cosmic microwave background $T_{\text{CMB}} = 2.7\text{ K}(1+z)$. At $z \approx 15$, this results in a minimum fragment mass of $M_* \approx 20M_{\odot}(n_f/10^4\text{ cm}^{-3})^{-1/2}$, where $n_f > 10^4\text{ cm}^{-3}$ is the density at which opacity prevents further fragmentation²⁷. It is possible that the transition from the high-mass to the low-mass star formation mode was modulated by the cosmic microwave background temperature and was therefore gradual⁶, involving intermediate-mass ('population II.5') stars¹⁰ at intermediate redshifts. This transitional population can give rise to the faint supernovae that have been proposed to explain the observed abundance patterns in metal-poor stars^{15,17}. These solar-mass stars themselves should have formed out of metal-poor gas at relatively late times when the cosmic microwave background temperature was sufficiently low.

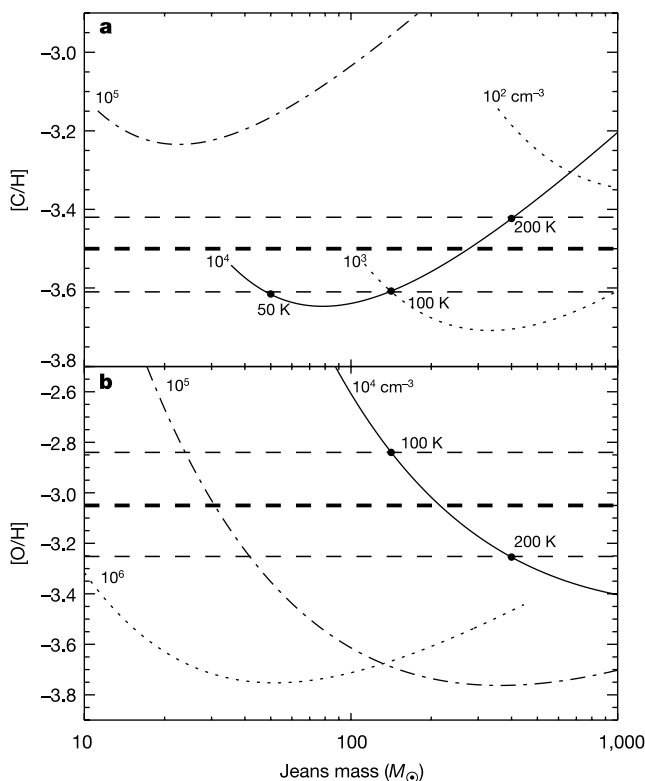


Figure 1 Required carbon and oxygen abundances (relative to solar values) for cooling a clump of metal-poor gas faster than its free-fall (compressional heating) time. The minimum abundances are shown as functions of the initial Jeans mass of the clump (with higher abundances allowing the clump to cool and further fragment). **a**, Singly ionized carbon; **b**, neutral oxygen. The curves are labelled by the hydrogen number density in units of cm^{-3} , with the solid curve marking the characteristic density in pre-stellar clumps of metal-poor gas, $n_c \approx 10^4\text{ cm}^{-3}$. Heavy filled dots indicate particular temperature values. The critical abundance is delineated by the horizontal dashed lines, with a heavy line indicating the central value and the adjacent light lines bracketing the uncertainty within the characteristic temperature range, $T_c \approx 100\text{--}200\text{ K}$, and density, $\sim 10^4\text{ cm}^{-3}$, of pre-stellar clumps. The figure indicates the minimum metal abundance at which fragmentation into low-mass stars starts, but it does not provide information about the minimum mass of the fragments at the end of the cooling process.

In relating our results to theoretical nucleosynthetic yields, one has to address the degree of mixing, or dilution, that the supernovae ejecta experience. Supernovae at high redshifts occur in dark matter halos with shallow potential wells, so that the resulting blast wave is predicted²⁸ to escape into the expanding intergalactic medium. The dilution mass can then be estimated as $M_{\text{dil}} \approx 2E_{\text{SN}}/v^2$, where E_{SN} is the explosion kinetic energy, and $v = H(z)r$ is the Hubble velocity at a distance r . Using standard cosmological parameters and ignoring radiative losses²⁸, we get $M_{\text{dil}} \approx 10^7 M_{\odot}(E_{\text{SN}}/10^{53}\text{ erg})^{3/5}[(1+z)/16]^{-3/5}$. In the expected mass range of the pair-instability supernovae¹² and the corresponding explosion energies $E_{\text{SN}} \approx 10^{51} - 10^{53}\text{ erg}$, one finds dilution masses that are more than two orders of magnitude larger than typical values for type II supernovae in nearby galaxies¹⁴. However, depending on the exact values of the ejected mass of C and O and the respective explosion energies, a single supernova event could already lead to [C/H] and [O/H] ratios that 'overshoot' the critical abundances. For example, the abundances resulting from the explosion and subsequent mixing of a pair-instability supernova at the low-mass end¹² are [C/H] ≈ -2.9 and [O/H] ≈ -2.4 , respectively, roughly 0.5 dex above the critical levels. Nevertheless, a single pair-instability supernova at the high-mass end would produce final, mixed abundances still below critical. Because the dilution mass scales as the number of simultaneous supernovae to the 3/5 power while the ejecta mass scales linearly with this number, multiple simultaneous supernovae

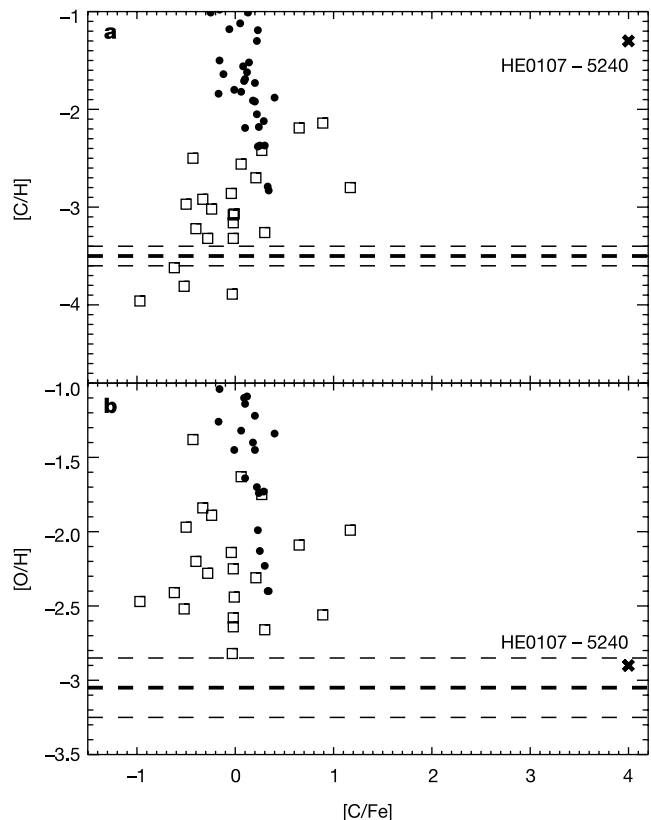


Figure 2 Observed abundances in low-metallicity Galactic halo stars. For both carbon (**a**) and oxygen (**b**), filled circles correspond to samples of dwarf and subgiant stars²⁹, and open squares to a sample of giant stars³⁰. Both data sets were obtained at a high signal-to-noise ratio using the ultraviolet-visual echelle spectrograph (UVES) on the Very Large Telescope (VLT). The dashed lines indicate the predicted level of critical carbon and oxygen abundances from Fig. 1. According to our theoretical model, all low-mass dwarf stars ought to lie above the critical threshold level of at least one of the two elements. Indeed all stars in the existing data sets satisfy this condition. We highlight the location of the extremely iron-poor giant star HE0107-5240 (marked by a cross), whose carbon and oxygen abundances are both above the critical levels⁴.

within the same galaxy would be less effective at diluting their metal abundances into the expanding intergalactic medium. But further dilution may occur during the hierarchical assembly of galaxies, as infall may add metal-free gas from the intergalactic medium to the enriched interstellar medium.

In Fig. 2 we compare our theoretical thresholds to the observed C and O abundances in metal-poor dwarf²⁹ and giant³⁰ stars in the halo of our Galaxy. As can be seen, all data points lie above the critical O abundance but a few cases lie below the critical C threshold. All of these low-mass stars are consistent with our model because the corresponding O abundances lie above the predicted threshold. The sub-critical [C/H] abundances could have originated either in the progenitor cloud or from the mixing of CNO-processed material (with carbon converted into nitrogen) into the stellar atmosphere during the red giant phase. To guard against this a posteriori processing of C, dwarf stars should be preferred, as they provide a more reliable record of the primordial abundance pattern that was present at birth. However, because dwarfs are fainter, another strategy would be to focus on giants which show evidence for weak mixing (for example, in terms of high ¹²C/¹³C isotope ratio or low ¹⁴N abundance). The current data on dwarf stars (filled symbols in Fig. 2) does not yet reach C and O abundances that are sufficiently low to probe the theoretical cooling thresholds. Note also that the extremely iron-poor star HE0107–5240 has C and O abundances that both lie above the respective critical levels (Bessell, M. *et al.*, manuscript in preparation). The formation of this low-mass star ($\sim 0.8M_{\odot}$) is therefore consistent with the theoretical framework considered here^{11,17}.

Our model suggests a powerful new probe of the first supernovae to have exploded in the Universe. By selecting an unbiased new sample of Galactic halo stars with C or O abundances close to the theoretical critical values, it should be possible to sample individual supernova events originating in the earliest epochs of cosmic star formation. The value of [C/Fe] would be indicative of the particular supernova type. For example, the very high value of [C/Fe] $\approx +4$ observed⁴ in HE0107–5240 can be explained by a core-collapse supernova which leaves behind a black hole after fallback of most of the outer stellar envelope¹⁷. Modern surveys select metal-poor stars based on their weak or absent Ca II K-lines. Our model suggests that a better strategy to select truly ‘second generation’ stars is to look for stars with the lowest abundances of carbon and oxygen. □

Received 18 August; accepted 18 September 2003; doi:10.1038/nature02071.

1. Abel, T., Bryan, G. & Norman, M. The formation of the first star in the universe. *Science* **295**, 93–98 (2002).
2. Bromm, V., Coppi, P. S. & Larson, R. B. The formation of the first stars. I. The primordial star-forming cloud. *Astrophys. J.* **564**, 23–51 (2002).
3. Nakamura, F. & Umemura, M. The stellar initial mass function in primordial galaxies. *Astrophys. J.* **569**, 549–557 (2002).
4. Christlieb, N. *et al.* A stellar relic from the early Milky Way. *Nature* **419**, 904–906 (2002).
5. McKee, C. F. & Tan, J. C. Massive star formation in 100,000 years from turbulent and pressurized molecular clouds. *Nature* **416**, 59–61 (2002).
6. Clarke, C. J. & Bromm, V. The characteristic stellar mass as a function of redshift. *Mon. Not. R. Astron. Soc.* **343**, 1224–1230 (2003).
7. Omukai, K. Protostellar collapse with various metallicities. *Astrophys. J.* **534**, 809–824 (2000).
8. Bromm, V., Ferrara, A., Coppi, P. S. & Larson, R. B. The fragmentation of pre-enriched primordial objects. *Mon. Not. R. Astron. Soc.* **328**, 969–976 (2001).
9. Schneider, R., Ferrara, A., Natarajan, P. & Omukai, K. First stars, very massive black holes, and metals. *Astrophys. J.* **571**, 30–39 (2002).
10. Mackey, J., Bromm, V. & Hernquist, L. Three epochs of star formation in the high-redshift universe. *Astrophys. J.* **586**, 1–11 (2003).
11. Schneider, R., Ferrara, A., Salvaterra, R., Omukai, K. & Bromm, V. Low-mass relics of early star formation. *Nature* **422**, 869–871 (2003).
12. Heger, A. & Woosley, S. E. The nucleosynthetic signature of population III. *Astrophys. J.* **567**, 532–543 (2002).
13. Qian, Y.-Z., Sargent, W. L. W. & Wasserburg, G. J. The prompt inventory from very massive stars and elemental abundances in Ly α systems. *Astrophys. J.* **569**, L61–L64 (2002).
14. Qian, Y.-Z. & Wasserburg, G. J. Determination of nucleosynthetic yields of supernovae and very massive stars from abundances in metal-poor stars. *Astrophys. J.* **567**, 515–531 (2002).
15. Umeda, H. & Nomoto, K. Nucleosynthesis of zinc and iron peak elements in population III type II supernovae. *Astrophys. J.* **565**, 385–404 (2002).
16. Sneden, C. & Cowan, J. J. Genesis of the heaviest elements in the Milky Way Galaxy. *Science* **299**, 70–75 (2003).

17. Umeda, H. & Nomoto, K. First-generation black-hole-forming supernovae and the metal abundance pattern of a very iron-poor star. *Nature* **422**, 871–873 (2003).
18. Hollenbach, D. & McKee, C. F. Molecule formation and infrared emission in fast interstellar shocks. *Astrophys. J.* **342**, 306–336 (1989).
19. Tumlinson, J. & Shull, J. M. Zero-metallicity stars and the effects of the first stars on reionization. *Astrophys. J.* **528**, L65–L68 (2000).
20. Bromm, V., Kudritzki, R. P. & Loeb, A. Generic spectrum and ionization efficiency of a heavy initial mass function for the first stars. *Astrophys. J.* **552**, 464–472 (2001).
21. Cen, R. The implications of Wilkinson Microwave Anisotropy Probe observations for population III star formation processes. *Astrophys. J.* **591**, L5–L8 (2003).
22. Sokasian, A., Yoshida, N., Abel, T., Hernquist, L. & Springel, V. Cosmic reionisation by stellar sources: Population III stars. *Mon. Not. R. Astron. Soc.* (submitted); preprint at (<http://xxx.lanl.gov/astro-ph/0302213>) (2003).
23. Wyithe, J. S. B. & Loeb, A. Was the universe reionized by massive metal-free stars? *Astrophys. J.* **588**, L69–L72 (2003).
24. Kogut, A. *et al.* Wilkinson Microwave Anisotropy Probe WMAP first year observations: TE polarization. *Astrophys. J. Suppl.* **148**, 161–173 (2003).
25. Bromm, V. & Loeb, A. Formation of the first supermassive black holes. *Astrophys. J.* **596**, 34–46 (2003).
26. Larson, R. B. Early star formation and the evolution of the stellar initial mass function in galaxies. *Mon. Not. R. Astron. Soc.* **301**, 569–581 (1998).
27. Rees, M. J. Opacity-limited hierarchical fragmentation and the masses of protostars. *Mon. Not. R. Astron. Soc.* **176**, 483–486 (1976).
28. Bromm, V., Yoshida, N. & Hernquist, L. The first supernova explosions in the universe. *Astrophys. J.* (in the press); preprint at (<http://xxx.lanl.gov/astro-ph/0305333>) (2003).
29. Akerman, C. J., Carigi, L., Nissen, P. E., Pettini, M. & Asplund, M. The evolution of the C/O ratio in metal-poor halo stars. *Astron. Astrophys.* (submitted).
30. Cayrel, R. *et al.* Abundance patterns and supernova yields in the early Galaxy from C to Zn. *Astron. Astrophys.* (submitted).

Acknowledgements We thank R. Barkana and particularly T. Beers for discussions, and are grateful to T. Beers and M. Pettini for making their data available prior to publication. This work was supported in part by NSF, NASA and the Guggenheim foundation (for A.L.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to V.B. (vbromm@cfa.harvard.edu) or A.L. (loeb@cfa.harvard.edu).

A coherent three-dimensional Fermi surface in a high-transition-temperature superconductor

N. E. Hussey¹, M. Abdel-Jawad¹, A. Carrington¹, A. P. Mackenzie² & L. Balicas³

¹H. H. Wills Physics Laboratory, University of Bristol, Tyndall Avenue, Bristol BS8 1TL, UK

²School of Physics and Astronomy, University of St Andrews, St Andrews, Fife KY16 9SS, UK

³National High Magnetic Field Laboratory, Florida State University, Tallahassee, Florida 32306, USA

All conventional metals are known to possess a three-dimensional Fermi surface, which is the locus in reciprocal space of the long-lived electronic excitations that govern their electronic properties at low temperatures. These excitations should have well-defined momenta with components in all three dimensions. The high-transition-temperature (high- T_c) copper oxide superconductors have unusual, highly two-dimensional properties above the superconducting transition¹. This, coupled with a lack of unambiguous evidence for a three-dimensional Fermi surface, has led to many new and exotic models for the underlying electronic ground state². Here we report the observation of polar angular magnetoresistance oscillations³ in the overdoped superconductor $Tl_2Ba_2CuO_{6+\delta}$ in high magnetic fields, which firmly establishes the existence of a coherent three-dimensional Fermi surface. Analysis of the oscillations reveals that at certain symmetry points, however, this surface is strictly two-dimen-

sional. This striking form of the Fermi surface topography, long-predicted by electronic band structure calculations⁴, provides a natural explanation for a wide range of anisotropic properties both in the normal^{5,6} and superconducting states^{7–9}. Our data reveal that, despite their extreme electrical anisotropy, the high- T_c materials at high doping levels can be understood within a framework of conventional three-dimensional metal physics.

A quasi-two-dimensional (Q2D) metal is defined as a metal with an open (that is, cylindrical) Fermi surface (FS) that has only small but finite warping along one of its principal reciprocal lattice vectors. In a strong magnetic field $\mathbf{H}$ (or more precisely, when the product of the cyclotron frequency and scattering time $\omega_c\tau \approx 1$), the interlayer resistivity $\rho_{\perp}$ of a Q2D metal oscillates as $\mathbf{H}$ is rotated in a polar plane relative to the conducting layers (that is, varying angle θ as defined in the left inset to Fig. 1)³. Maxima in $\rho_{\perp}$ occur when the velocity perpendicular to the layers $v_{\perp}$, averaged over many trajectories on the FS, is zero. These polar angular magnetoresistance oscillations (AMRO) have proved to be an invaluable probe of FS topography in many low-dimensional conductors.

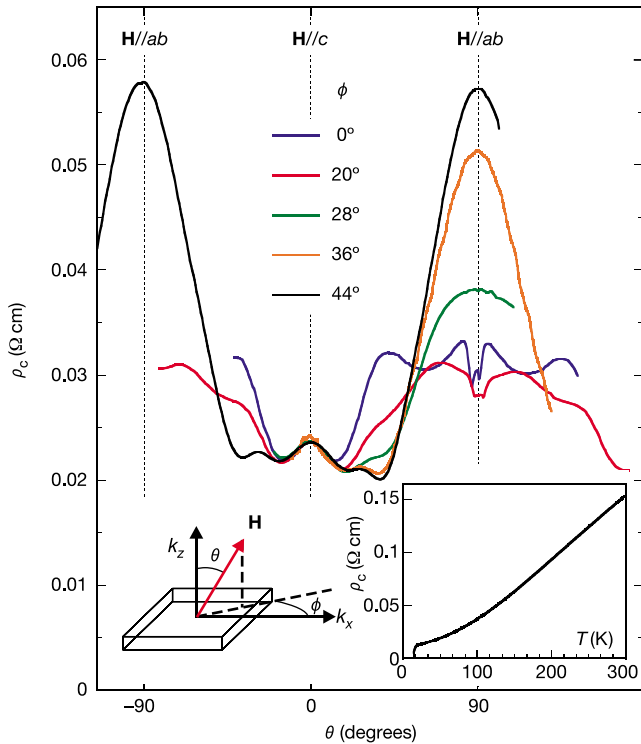


Figure 1 Polar AMRO sweeps in an overdoped Tl2201 single crystal ($T_c \approx 20$ K). The data were taken at $T = 4.2$ K and $H = 45$ T. The different azimuthal orientations ($\pm 4^\circ$) of each polar sweep are stated relative to the Cu–O–Cu bond direction. The key features of the data are as follows: (1) a sharp dip in $\rho_{\perp}$ at $\theta = 90^\circ$ for low values of ϕ , which we attribute to the onset of superconductivity at angles where $H_{c2}(\phi, \theta)$ is maximal, (2) a broad peak around $\mathbf{H} \parallel \mathbf{ab}$ ($\theta = 90^\circ$) that is maximal for $\phi \approx 45^\circ$, consistent with previous azimuthal AMRO studies in overdoped Tl2201 (ref. 16), (3) a small peak at $\mathbf{H} \parallel \mathbf{c}$ ($\theta = 0^\circ$), and (4) a second peak in the range $25^\circ < \theta < 45^\circ$ whose position and intensity vary strongly with ϕ . These last two features are the most critical for our analysis. Similar features were observed in data from five other crystals. The slight asymmetry in $\Delta\rho_{\perp}(\theta)$ either side of $\mathbf{H} \parallel \mathbf{c}$ is believed to arise from a small Hall component in the measured signal. Inset, zero-field c -axis resistivity $\rho_c(T)$ of the same crystal. Note the good metallicity of $\rho_c(T)$, despite the large resistive anisotropy $\rho_{\perp}/\rho_{ab} \approx 1,000$ (ref. 16). Tl2201 crystals (typical dimensions $0.2 \times 0.1 \times 0.02$ mm³) were grown via a self-flux method as described previously¹³. Crystals selected from a particular batch were annealed in flowing oxygen at 360°C for one week in order to achieve the appropriate doping level before measuring their zero-field $\rho_{\perp}(T)$ with a four-terminal a.c. method. The high-field magnetoresistance measurements were carried out using a single-axis rotator and a ^3He cryostat inserted into a 45 T hybrid magnet.

Importantly, detailed comparisons made on Sr_2RuO_4 (ref. 30) and the organic conductors^{10,11} have shown that both the size and shape of the FS extracted from such AMRO measurements is consistent with that determined by more rigorous quantum oscillation studies.

The applicability of AMRO to the high- T_c copper oxides has been hampered by their large normal-state scattering rates and high upper critical fields H_{c2} . Monolayer $\text{Tl}_2\text{Ba}_2\text{CuO}_{6+\delta}$ (Tl2201), however, has a number of favourable characteristics that make it a suitable candidate for polar AMRO studies. Firstly, it is relatively clean, with low residual (in-plane) resistivities $\rho_{ab}(0) < 10 \mu\Omega\text{cm}$ (refs 12–15) (corresponding to $\omega_c\tau > 0.5$ at 60 T (ref. 14)). Secondly, as shown in the right-hand inset to Fig. 1, $\rho_{\perp}(T)$ in overdoped Tl2201 shows good metallic behaviour¹⁶. Overdoped Tl2201 also has a low H_{c2} (ref. 15). Finally, with respect to the interpretation of AMRO data¹⁷, electronic band structure calculations predict a relatively simple single-band FS taking the form of a slightly warped hole-like cylinder^{17,18}.

Figure 1 shows AMRO data $\Delta\rho_{\perp}(\phi, \theta)$ for an overdoped Tl2201 single crystal ($T_c \approx 20$ K) taken at 4.2 K and 45 T. The different coloured traces represent polar AMRO sweeps at given azimuthal angles ϕ relative to the Cu–O–Cu bond direction ($\{k_x, k_y\} = (\pi, 0)$). Each sweep contains significant θ -dependent structure that is summarized in the figure legend. We believe such strong θ - and ϕ -dependence can only arise from a coherent Q2D FS with significant in-plane anisotropy. Such a Q2D FS is most elegantly expressed¹⁹ in the form of an expansion:

$$k_F(\phi, \kappa) = \sum_{\substack{m, n=0 \\ n \text{ even}}} k_{mn} \cos n\kappa \begin{cases} \cos m\phi & (m \bmod 4 = 0) \\ \sin m\phi & (m \bmod 4 = 2) \end{cases} \quad (1)$$

where $\kappa = k_z d$ and d is the interlayer spacing ($=1.16$ nm for Tl2201). Before proceeding with the analysis, however, it is important to respect certain symmetry constraints on the general expansion for a hole-like FS closed around the X point of the Brillouin zone. Firstly, the body-centred-tetragonal (b.c.t.) symmetry of the Tl2201 crystal lattice imposes the condition that coefficients k_{mn} are non-zero only for n even and m divisible by 4, or for n odd and $m \bmod 4 = 2$ (ref. 19). Secondly, the role of oxygen bonding orbitals in mediating the interplane hopping means that the dominant hopping is between adjacent planes (that is, $n = 1$). Following the standard procedure for a series expansion, we therefore proceed by expanding to the lowest order necessary to satisfy these constraints and fit to the data. In the current simple model, this requires retaining terms k_{00} , k_{40} , k_{21} , and possibly higher-order terms in the $n = 1$ series. $\omega_c\tau$ is another free parameter that sets the magnitude, but not the angular dependence, of the oscillations.

Figure 2a shows the experimental data from Fig. 1 re-plotted for $0^\circ \leq \theta \leq 80^\circ$. In order to simulate the experimental data, we calculated $\Delta\rho_{\perp}(\phi, \theta)$ using the Shockley–Chambers tube-integral form of the Boltzmann transport equation²⁰ modified for a Q2D metal. The resultant simulations for different combinations of the appropriate harmonic terms in $k_F(\phi, \kappa)$ are shown in Fig. 2b–d. The combination k_{00} , k_{40} and k_{21} (Fig. 2b) is found to give strong ϕ -dependence in $\Delta\rho_{\perp}(\theta)$ and a peak at $\mathbf{B} \parallel \mathbf{c}$, but fails to reproduce the data satisfactorily at large angles. Adding the next term k_{61} (Fig. 2c) substantially improves the fit, but by going one order further to include k_{101} (Fig. 2d), we obtain excellent agreement with the experimental data. A single value of $\omega_c\tau$ ($= 0.45$) was found to best match the overall scale of the AMRO, and for simplicity this was used for all simulations shown. As a consistency check on our fitting procedure, $\omega_c\tau$ was estimated independently from in-plane resistivity ρ_{ab} and Hall effect R_H measurements on similarly doped crystals^{12,13}. Given $R_H \approx 8(1) \times 10^{-8} \text{ m}^3 \text{ C}^{-1}$ and $\rho_{ab} \approx 8(2) \mu\Omega\text{cm}$, we find $\omega_c\tau = H\sigma_{xy}/\sigma_{xx} = HR_H/\rho_{ab} \approx 0.45$ (0.1) in 45 T, in agreement with our AMRO analysis. Thus, we believe our final

expression for $k_F(\phi, \kappa)$, displayed in the legend for Fig. 2d, represents a robust empirical determination of the FS dispersion of Tl2201.

The overall shape of the FS is consistent with features of band structure calculations for optimally doped Tl2201 (ref. 18) and other copper oxides²¹ using the local density approximation. The size of the derived FS occupies around 62% of the first Brillouin zone, equivalent to a hole doping per Cu atom of $p = 0.24$ and in good agreement with conclusions of a previous low-temperature Hall effect study¹². The particular form of $k_F(\phi)$ is consistent with the tetragonal symmetry of Tl2201, and gives the FS the appearance of a slightly rounded square within the ab -plane (see Fig. 3).

But the detail of our fit and associated parameters is less important than the principle. Whilst the observation of polar AMRO by itself is not proof of interlayer coherence in a Q2D metal²², the peak observed at $\mathbf{H} \parallel c$ coupled with the remarkable

sensitivity of our analysis to fine details in the c -axis warping provides compelling evidence that c -axis normal-state transport in overdoped Tl2201 is indeed coherent. Throughout nearly two decades of research on the copper oxides, the issue of coherence has been central to elucidating the dimensionality of the electronic ground state. Whilst evidence for the development of three-dimensionality has been inferred from previous measurements on overdoped cuprates, the data presented here are, to our knowledge, the first direct experimental proof that a three-dimensionally coherent FS exists in the normal state of any superconducting copper oxide. It should be noted that the observation of similar T -dependencies in $\rho_{\perp}(T)$ and $\rho_{ab}(T)$ is not sufficient to prove the existence of a 3D FS, since it has been shown theoretically²³ that the behaviour of $\rho_{\perp}(T)$ can mirror $\rho_{ab}(T)$ even if c -axis transport is incoherent. Furthermore, evidence for c -axis coherence from optical conductivity has been limited to heavily -overdoped non-superconducting copper oxides²⁴. Finally, although angle-resolved photoemission spectroscopy (ARPES) has provided detailed information about the topology of the in-plane FS²⁵, it has yet to resolve any finite interlayer dispersion, apart from that due to bilayer splitting in $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_8$.

Figure 3 shows our empirically determined Tl2201 FS projected onto the ab -plane in order to highlight the in-plane variation of the c -axis hopping integral $t_{\perp}(\phi)$. Significantly, c -axis dispersion is found to vanish along eight specific symmetry lines of the Brillouin zone, in the so-called nodal directions along (π, π) (for example, ΓX in Fig. 3), where the in-plane quasiparticles are longest-lived, and also in the anti-nodal regions $(\pi, 0)$, where in-plane scattering is most intense and quasiparticles (at optimal doping and below) are incoherent above T_c (refs 5, 25, 26).

Recognition of the role of oxygen bonding and virtual Cu 4s orbitals in c -axis conduction led to a well-known prediction^{4,7} of the existence of four dispersive nodes (that is, zeros in $t_{\perp}(\phi)$) in the simple-tetragonal (Y- and Hg-based) copper oxides (centred around $(\pm\pi, \pm\pi)$) and eight zeros in the b.c.t. (La-, Bi- and Tl-based) copper oxides as found here. They have formed an integral

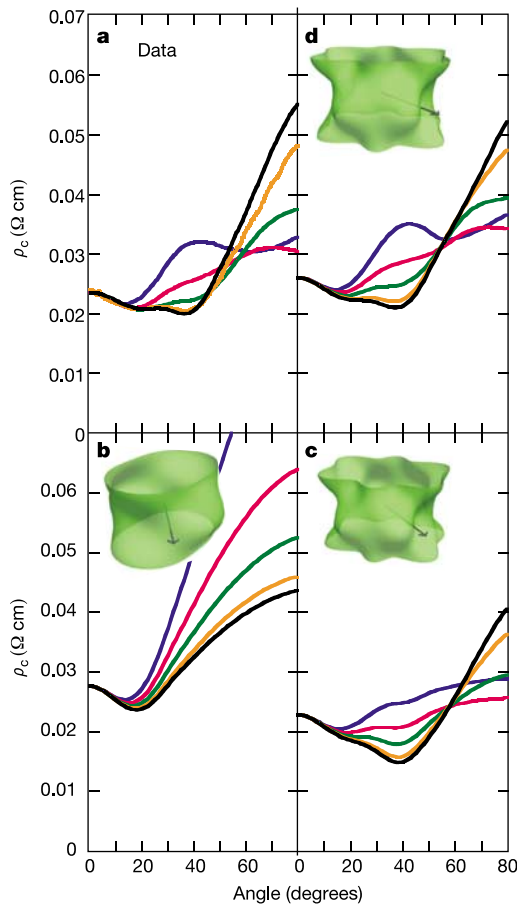


Figure 2 Reconstruction of the FS in Tl2201 from polar AMRO data. **a**, The experimental data used to generate the FS are the same as those in Fig. 1 plotted over the range $0^\circ \leq \theta \leq 80^\circ$. **b**, Simulated AMRO data for a Q2D FS (shown schematically in relief, the arrow indicating the direction of the Cu–O bond direction) with $k_F(\phi, \kappa) = k_{00} + k_{40} \cos 4\phi + k_{21} \cos \kappa \sin 2\phi$. **c**, As **b** but with $k_F(\phi, \kappa) = k_{00} + k_{40} \cos 4\phi + \cos \kappa (k_{21} \sin 2\phi + k_{61} \sin 6\phi)$. **d**, As **c** but with $k_F(\phi, \kappa) = k_{00} + k_{40} \cos 4\phi + \cos \kappa (k_{21} \sin 2\phi + k_{61} \sin 6\phi + k_{101} \sin 10\phi)$. For all simulations, $\omega_c \tau = 0.45$, $k_{00} = 7.45$, $k_{40} = -0.19$, $k_{21} = 0.31$, $k_{61} = 0.021$ and $k_{101} = -0.0085$. All k_{mn} coefficients are expressed in units of nm^{-1} . The absolute values of k_{21} , k_{61} and k_{101} were estimated from the resistive anisotropy ($\rho_{\perp}/\rho_{ab} \approx 10^3$) assuming isotropic τ and have associated error bars of order $\pm 10\%$. A previous theoretical paper on AMRO in Tl2201 (ref. 17) predicted different polar patterns to those observed here, the most significant difference being an absence of the AMRO peak at $\theta = 0$ ($\mathbf{H} \parallel c$). Their calculation, however, used an oversimplified $t_{\perp}(\phi)$ that did not contain any nodes. It is important to realize that the appearance of a peak at $\theta = 0$ implies that $v_{\perp}$ averages to zero even for strictly in-plane cyclotron orbits and this can only occur with the inclusion of components k_{21} , k_{61} , and so on, in the FS expansion.

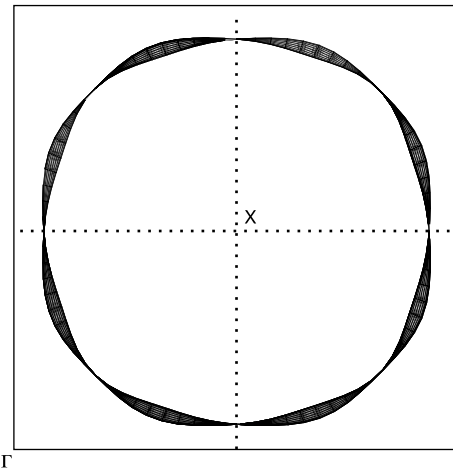


Figure 3 Projection of the deduced FS of Tl2201 onto the ab -plane. The magnitude of the c -axis warping has been increased fourfold to emphasize the eight loci where k_z dispersion vanishes. The Brillouin zone has been simplified from its usual b.c.t. form so as to accommodate the full ab -plane projection. The experimental verification of the condition for zeros in $t_{\perp}(\phi)$, namely $k_{21} - k_{61} + k_{101} \approx 0$, is confirmed by inspection of the fitting parameters in Fig. 2. We note that original band calculations on optimally doped Tl2201 (ref. 18) also predicted a small elliptical FS sheet due to the Tl–O layers. We emphasize that neither this work nor any of the other magnetotransport studies on Tl2201 (refs 12–14, 16) gives any evidence for its existence. In particular, the observed resistive anisotropy¹⁶ is orders of magnitude too large to support a three-dimensional (3D) FS sheet of the type predicted.

part of numerous theoretical treatments of the physics of high-temperature superconductors, in both the normal and superconducting states, but until now their presence had never been directly confirmed experimentally. This locally non-dispersive structure has been invoked, for example, to explain the striking differences observed between the in-plane and out-of-plane normal state conductivities $\sigma_{ab}(\omega, T)$ and $\sigma_{\perp}(\omega, T)$ (refs 5, 6). In the superconducting state, these $t_{\perp}(\phi)$ zeros coincide precisely with nodes in the d -wave order parameter²⁷. This vanishing of c -axis tunnelling right at the gap nodes is thought to be responsible for the striking differences in the T -dependencies of the in-plane and out-of-plane penetration depths $\lambda_{ab}(T)$ and $\lambda_{\perp}(T)$ (ref. 7), the absence of the zero-bias conductance peak inside the d -wave vortex core⁸ and the slowly decaying tails in the tunnelling conductance around each vortex⁹.

Finally, the experimental demonstration of the existence of a 3D coherent FS in Tl2201 represents a significant challenge to models of high- T_c superconductivity based on purely 2D electron motion within the CuO₂ planes. Although Tl2201 at this level of doping clearly comes from the side of the phase diagram in which more conventional behaviour is observed²⁸, it is still a high-temperature superconductor ($T_c \approx 20$ K). Moreover, its normal state resistive anisotropy ($\rho_{\perp}/\rho_{ab} \approx 1,000$; ref. 16) is actually larger than that seen in some optimally doped materials, so it seems likely that the findings reported here will apply to a wider range of dopings. In common with other studies of Tl2201 (ref. 29), this work supports a model of the normal state based around fermionic quasiparticles. How and where this model breaks down as the Mott insulator is approached at half-filling of course remains an important outstanding issue. However, in providing crucial information on the dimensionality of the conducting planes, AMRO may prove to become an ideal complementary probe to ARPES in exploring this evolution of the electronic ground state in copper oxides across the phase diagram. □

Received 7 May; accepted 7 August 2003; doi:10.1038/nature01981.

- Ong, N. P. Non-Fermi-liquid aspects of charge-transport in cuprate superconductors. *Physica C* **235**, 221–224 (1994).
- Orenstein, J. & Millis, A. J. Advances in the physics of high temperature superconductivity. *Science* **288**, 468–474 (2000).
- Yamaji, K. On the angle dependence of the magnetoresistance in quasi-two-dimensional organic superconductors. *J. Phys. Soc. Jpn* **58**, 1520–1522 (1989).
- Andersen, O. K., Liechtenstein, A. I., Jepsen, O. & Paulsen, F. LDA energy bands, low-energy Hamiltonians t' , t'' , $t_{\perp}(k)$ and $J_{\perp}$. *J. Phys. Chem. Solids* **56**, 1573–1591 (1995).
- Ioffe, L. B. & Millis, A. J. Zone-diagonal-dominated transport in high- T_c cuprates. *Phys. Rev. B* **58**, 11631–11637 (1998).
- van der Marel, D. Anisotropy of the optical conductivity of high- T_c cuprates. *Phys. Rev. B* **60**, R765–R769 (1999).
- Xiang, T. & Wheatley, J. M. c -axis superfluid response of copper oxide superconductors. *Phys. Rev. Lett.* **76**, 4632–4635 (1996).
- Wu, C., Xiang, T. & Su, Z.-B. Absence of the zero bias peak in vortex tunneling spectra of high-temperature superconductors. *Phys. Rev. B* **62**, 14427–14430 (2000).
- Franz, M. & Tesanovic, Z. Quasiparticle spectra in the vicinity of a d -wave vortex. *Phys. Rev. B* **60**, 3581–3588 (1999).
- Kartsovnik, M. V., Laukhin, V. N., Pesotskii, S. I., Schegolev, I. F. & Yakovenko, V. M. Angular magnetoresistance oscillations and the shape of the Fermi surface in β -(ET)₂IBr₂. *J. Phys. I* **2**, 89–99 (1992).
- Wosnitzer, J. et al. The Fermi surfaces of β -(BEDT-TTF)₂X. *J. Phys. I* **6**, 1597–1608 (1996).
- Mackenzie, A. P., Julian, S. R., Sinclair, D. C. & Lin, C. T. Normal state magnetotransport in superconducting Tl₂Ba₂CuO_{6+δ} to millikelvin temperatures. *Phys. Rev. B* **53**, 5848–5854 (1996).
- Tyler, A. W. *An Investigation into the Magnetotransport Properties of Layered Superconducting Perovskites*. Thesis, Univ. Cambridge (1997).
- Tyler, A. W. et al. High field study of normal state magnetotransport in Tl₂Ba₂CuO_{6+δ}. *Phys. Rev. B* **57**, R728–R731 (1998).
- Mackenzie, A. P. et al. Resistance upper critical field in Tl₂Ba₂CuO₆ at low temperatures and high magnetic fields. *Phys. Rev. Lett.* **71**, 1238–1241 (1993).
- Hussey, N. E. et al. Angular dependence of the c -axis normal state magnetoresistance in single crystal Tl₂Ba₂CuO_{6+δ}. *Phys. Rev. Lett.* **76**, 122–125 (1996).
- Dragulescu, A., Yakovenko, V. M. & Singh, D. J. Theory of angular magnetoresistance oscillations in Tl₂Ba₂CuO₆. *Phys. Rev. B* **60**, 6312–6315 (1999).
- Singh, D. J. & Pickett, W. E. Electronic characteristics of Tl₂Ba₂CuO₆. *Physica C* **203**, 193–199 (1992).
- Bergemann, C., Julian, S. R., Mackenzie, A. P., NishiZaki, S. & Maeno, Y. Detailed topography of the Fermi surface of Sr₂RuO₄. *Phys. Rev. Lett.* **84**, 2662–2665 (2000).
- Ziman, J. M. *Principles of the Theory of Solids* (Cambridge Univ. Press, Cambridge, 1972).
- Pavarini, E., Dasgupta, I., Saha-Dasgupta, T., Jepsen, O. & Andersen, O. K. Band-structure trend in hole-doped cuprates and correlation with T_{max} . *Phys. Rev. Lett.* **87**, 047003 (2001).

- McKenzie, R. H. & Moses, P. Incoherent interlayer transport and angular-dependent magnetoresistance oscillations in layered metals. *Phys. Rev. Lett.* **81**, 4492–4495 (1998).
- Kumar, N. & Jayannavar, A. M. Temperature dependence of the c -axis resistivity of high- T_c layered oxides. *Phys. Rev. B* **45**, 5001–5004 (1992).
- Tamasaku, K., Ito, T., Takagi, H. & Uchida, S. Interplane charge dynamics in La_{2-x}Sr_xCuO₄. *Phys. Rev. Lett.* **72**, 3088–3091 (1994).
- Damascelli, A., Hussain, Z. & Shen, Z.-X. Angle-resolved photoemission studies of the cuprate superconductors. *Rev. Mod. Phys.* **75**, 473–541 (2003).
- Hussey, N. E. The normal state scattering rate in high- T_c cuprates. *Eur. Phys. J. B* **31**, 495–507 (2003).
- Hussey, N. E. Low-energy quasi-particles in high- T_c cuprates. *Adv. Phys.* **51**, 1685–1771 (2002).
- Nakamae, S. et al. Electronic ground state of heavily overdoped non-superconducting La_{2-x}Sr_xCuO₄. *Phys. Rev. B* **68**, 100502 (2003).
- Proust, C., Boaknin, E., Hill, R. W., Taillefer, L. & Mackenzie, A. P. Heat transport in a strongly overdoped cuprate: Fermi liquid and a pure d -wave BCS superconductor. *Phys. Rev. Lett.* **89**, 147003 (2002).
- Bergemann, C. et al. Quasi-two-dimensional Fermi liquid properties of the unconventional superconductor Sr₂RuO₄. *Adv. Phys.* (in the press).

Acknowledgements We acknowledge technical support from S. Hannahs, and discussions with O. K. Andersen, C. Bergemann, S. J. Blundell, B. L. Gyorrfy, R. Hlubina, A. J. Leggett, G. Santi and J. A. Wilson. We also acknowledge J. R. Cooper for first recognizing the potential of AMRO studies in Tl2201. This work was supported by EPSRC, the Leverhulme Trust and by a cooperative agreement between the State of Florida and NSF. The crystals were grown by A.P.M., the experiments carried out by M.A.-J., L.B., A.C. and N.E.H., and the analysis performed by M.A.-J. and N.E.H. in discussion with A.P.M.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to N.E.H. (n.e.hussey@bristol.ac.uk).

A broadband superconducting detector suitable for use in large arrays

Peter K. Day¹, Henry G. LeDuc¹, Benjamin A. Mazin², Anastasios Vayonakis² & Jonas Zmuidzinas²

¹Jet Propulsion Laboratory, Pasadena, California 91107, USA

²California Institute of Technology, 320-47, Pasadena, California 91125, USA

Cryogenic detectors are extremely sensitive and have a wide variety of applications^{1–3} (particularly in astronomy^{4–8}), but are difficult to integrate into large arrays like a modern CCD (charge-coupled device) camera. As current detectors of the cosmic microwave background (CMB) already have sensitivities comparable to the noise arising from the random arrival of CMB photons, the further gains in sensitivity needed to probe the very early Universe will have to arise from large arrays. A similar situation is encountered at other wavelengths. Single-pixel X-ray detectors now have a resolving power of $\Delta E < 5$ eV for single 6-keV photons, and future X-ray astronomy missions⁷ anticipate the need for 1,000-pixel arrays. Here we report the demonstration of a superconducting detector that is easily fabricated and can readily be incorporated into such an array. Its sensitivity is already within an order of magnitude of that needed for CMB observations, and its energy resolution is similarly close to the targets required for future X-ray astronomy missions.

Among cryogenic detectors, superconducting detectors are especially attractive because thin-film deposition and lithographic patterning techniques may be used to produce large arrays. Several types have been demonstrated, including tunnel junction detectors^{4,9,10} and transition-edge sensors¹¹. These have nearly achieved the desired performance^{5,6,8,12–14} for energy-resolved detection of individual optical/ultraviolet/X-ray photons and sensitive power detection at millimetre and submillimetre wavelengths. However,

these detectors have generally been used with individual preamplifiers and wiring for the output signals, which is clearly impractical for large arrays. Instead, a multiplexed readout approach is needed, in which preamplifiers and signal wiring are shared among multiple detectors. Multiplexing schemes are now being developed for transition-edge sensors^{15,16}, but will require complex, custom-designed superconducting electronics, located close to the detector array. Our detector concept¹⁷ is based on the microwave measurement of the complex impedance of a thin superconducting film, and allows a simple frequency-domain approach to multiplexing. This results in a dramatic simplification of the detector array and associated cryogenic electronics, and harnesses the rapid advances in wireless communications electronics. The results we present include the demonstration of single X-ray photon detection with a high signal-to-noise ratio and a measurement of the detector noise. Although much work remains to be done to optimize the performance, and to produce and use practical detector arrays, the devices already achieve very interesting levels of sensitivity.

In order to explain the operation of our detector, we must quickly review the electrodynamics of superconductors¹⁸. As its name implies, a superconductor has zero resistance for d.c. electrical

current. This supercurrent is carried by pairs of electrons, known as Cooper pairs. Cooper pairs are bound together by the electron-phonon interaction, with a binding energy $2\Delta \approx 3.5k_B T_c$, where T_c is the superconducting transition temperature. However, superconductors have a nonzero impedance for a.c. currents. An electric field applied near the surface of a superconductor causes the Cooper pairs to accelerate, allowing energy storage in the form of kinetic energy. Because the supercurrent is non-dissipative, this energy may be extracted by reversing the electric field. Similarly, energy may be stored in the magnetic field inside the superconductor, which penetrates only a short distance, $\lambda \approx 50$ nm, from the surface. The overall effect is that a superconductor has a surface inductance $L_s = \mu_0 \lambda$, due to the reactive energy flow between the superconductor and the electromagnetic field. The surface impedance $Z_s = R_s + i\omega L_s$ also includes a surface resistance R_s , which describes a.c. losses at angular frequency ω caused by the small fraction of electrons that are not in Cooper pairs, which are called 'quasiparticles'. For temperatures T much lower than T_c , $R_s \ll \omega L_s$.

Photons with sufficient energy ($h\nu > 2\Delta$) may break apart one or more Cooper pairs (Fig. 1a). The absorption of a high-energy photon creates $N_{qp} \approx \eta h\nu / \Delta$ quasiparticles; the excess quasiparti-

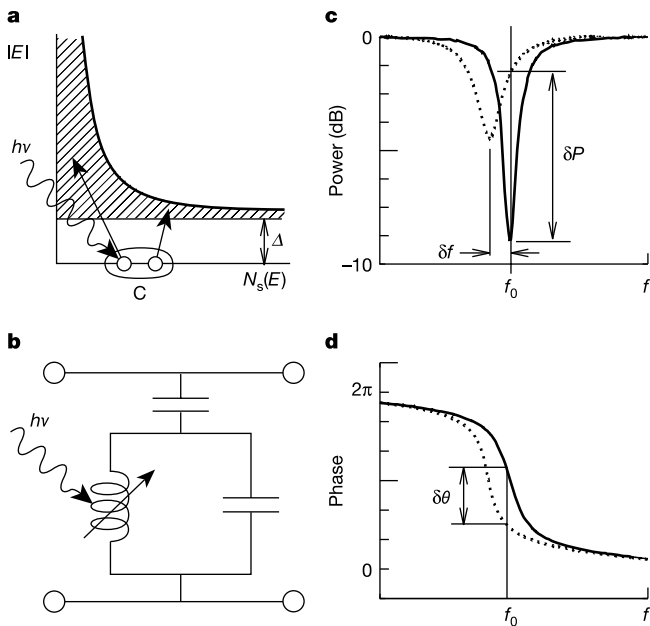


Figure 1 An illustration of the detection principle. **a**, Photons with energy $h\nu > 2\Delta$ are absorbed in a superconducting film cooled to $T \ll T_c$, breaking Cooper pairs and creating a number of quasiparticle excitations $N_{qp} = \eta h\nu / \Delta$. In this diagram, Cooper pairs (C) are shown at the Fermi level, and the density of states for quasiparticles¹⁸, $N_s(E)$, is plotted as the shaded area as a function of quasiparticle energy E . **b**, The increase in quasiparticle density changes the (mainly inductive) surface impedance $Z_s = R_s + i\omega L_s$ of the film, which is used as part of a microwave resonant circuit. The resonant circuit is depicted schematically here as a parallel LC circuit which is capacitively coupled to a through line. The effect of the surface inductance L_s is to increase the total inductance L , while the effect of the surface resistance R_s is to make the inductor slightly lossy (adding a series resistance). **c**, On resonance, the LC circuit loads the through line, producing a dip in its transmission. The quasiparticles produced by the photons increase both L_s and R_s , which moves the resonance to lower frequency (due to L_s), and makes the dip broader and shallower (due to R_s). Both of these effects contribute to changing the amplitude (producing power change δP) (**c**) and phase (**d**) of a microwave probe signal transmitted through the circuit. The definition of the phase angle used here is explained in Fig. 3. The amplitude and phase curves shown in this illustration are actually the data measured for the test device (Fig. 2) at 120 mK (solid lines) and 260 mK (dashed lines). This choice of circuit design, which has high transmission away from resonance, is very well suited for frequency-domain multiplexing, because multiple resonators operating at slightly different frequencies could all be coupled to the same through line.

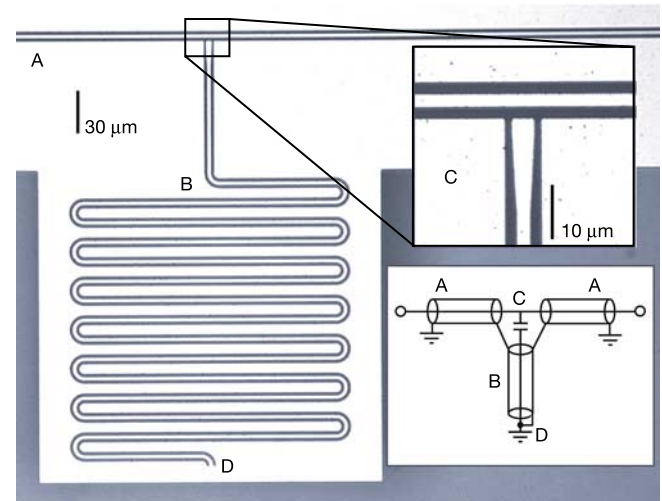


Figure 2 A microscope photograph of the device tested. Light and dark regions are the aluminium film and bare sapphire substrate, respectively. **A**, coplanar waveguide (CPW) through line used for excitation and readout. **B**, Meandered quarter-wavelength resonator section, with an overall length of 3 mm, and resonance frequency around 10 GHz. **C**, coupling capacitor. **D**, short-circuit termination. The coupling region is magnified in the inset; the diagram shows the equivalent circuit. Both CPW lines have a 50Ω characteristic impedance, and are fabricated from a single 2,200-Å-thick aluminium film ($T_c = 1.23$ K) using standard contact photolithography. The centre conductor of width $3 \mu\text{m}$ is separated by $2\text{-}\mu\text{m}$ gaps from ground planes on either side. The fraction α of the total inductance per unit length contributed by the surface inductance of the aluminium film can be written as a sum of centre strip and ground plane terms, $\alpha = \alpha_{\text{centre}} + \alpha_{\text{ground}}$. These are calculated to be $\alpha_{\text{centre}} \approx 0.04$ and $\alpha_{\text{ground}} \approx 0.02$ using a finite-element method²⁶, assuming an effective penetration depth $\lambda = 50$ nm. The measured resonator quality factor $Q = f_0 / \Delta f$ is 52,500 at low temperatures $T \ll T_c$ (Fig. 1). This device is mainly sensitive to photon events in the centre strip ($V = 2,000 \mu\text{m}^3$) of the CPW line rather than in the ground plane, because the microwave current in the ground plane is concentrated near the edge of the CPW line; quasiparticles generated in the ground plane near the edge of the CPW line can easily diffuse away. Similarly, the device is more sensitive to centre strip events occurring near the short-circuited end, where the standing-wave pattern of the microwave current reaches a maximum. Quasiparticles generated in the centre strip may also diffuse out of the short-circuited end; the peak response therefore occurs roughly one diffusion length (~ 1 mm) from this end. Photon events in the through line are not seen, because there is no resonant enhancement of the surface impedance effect.

cles subsequently recombine into Cooper pairs on timescales $\tau_{qp} \approx 10^{-3}$ – 10^{-6} s. Here $\eta \approx 0.57$ is the efficiency with which the photon energy is converted to quasiparticles¹⁹. During this time, the quasiparticles can diffuse over a distance of $l \approx \sqrt{D\tau_{qp}}$, where D is the diffusion constant of the material. Similarly, the absorption of a steady stream of low-energy (millimetre/submillimetre) photons would raise the steady-state quasiparticle density by an amount δn_{qp} above its thermal equilibrium value. Our detectors make use of the dependence of the surface impedance Z_s on quasiparticle density. Although the changes δZ_s are quite small, very sensitive measurements may be made using a resonant circuit (Fig. 1b). Changes in L_s and R_s affect the frequency and width of the resonance, respectively, changing the amplitude and phase of a microwave signal transmitted through the circuit (Fig. 1c and d).

The change in surface impedance δZ_s may be expressed as $\delta Z_s = \delta n_{qp} \partial Z_s / \partial n_{qp}$, where the derivative depends on ω , T , and material parameters, and may be calculated using the Mattis–Bardeen theory²⁰. Very crudely, we expect the fractional surface impedance change to be comparable to the fraction of Cooper pairs that are broken. This leads to the estimate $\delta L_s / L_s \approx \delta n_{qp} / 2N_0 \Delta$, where N_0 is the single spin density of states at the Fermi level, as $N_0 \Delta$ can be thought of as the density of Cooper pairs. The resulting fractional change in the resonance frequency f_0 is $|\delta f|/f_0 = 0.5\alpha \delta L_s / L_s$, where

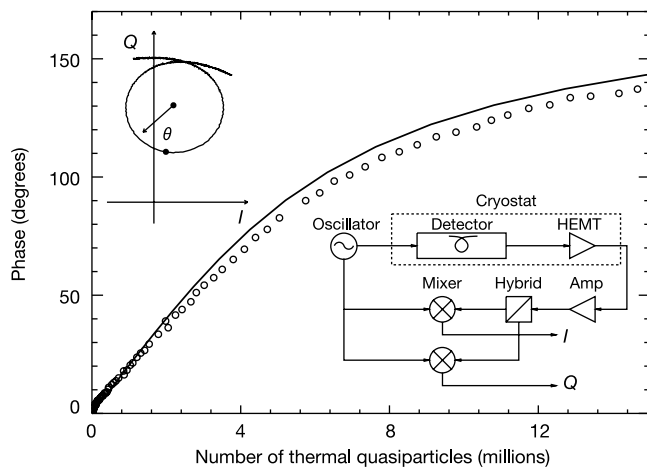


Figure 3 The microwave measurement set-up and the phase calibration of the detector. A low-phase noise microwave synthesizer (right inset) generates the fixed-frequency signal used to excite the detector, which is cooled in a dilution refrigerator. After transmission through the device, the signal is amplified by a broadband cryogenic HEMT amplifier with ~ 50 K noise temperature (including cable loss) and a room-temperature amplifier. An I/Q (inphase-quadrature) mixer is used to measure the amplitude and phase of the transmitted signal as a function of time. The left inset shows the measured resonance trajectory as a function of frequency in the I/Q plane (in arbitrary units), at 120 mK. The point plotted on the resonance circle indicates the resonance frequency. As shown by the arrow, the phase angle, θ , is calculated relative to this point using the centre of the circle as the reference. At low quasiparticle densities, this phase angle depends only on L_s and is proportional to the actual microwave phase shift through the device. The points on the main plot show the measured phase angle, θ , as a function of the calculated number of thermal quasiparticles in the central conductor of the resonator, $N_{qp}(T) = n_{qp}V = 2N_0V(2\pi k_B T \Delta)^{1/2} \exp(-\Delta/k_B T)$, obtained by varying the temperature T . Here $N_0 = 1.72 \times 10^{10} \mu\text{m}^{-3} \text{eV}^{-1}$ is the single-spin density of states at the Fermi surface, including the electron–phonon enhancement factor²⁷. The line gives the phase calculated from theory for $\alpha = 0.061$ (Fig. 2), $\Delta = 0.171$ meV, $Q(T=0) = 52,500$ (limited by coupling), and assuming the Mattis–Bardeen theory²⁰ for Z_s in the extreme anomalous limit (as $\xi_0 \approx 1,600$ nm $\gg \lambda_0 = 16$ nm for aluminium). The value of Δ is in reasonable agreement²⁸ with the measured $T_c = 1.23$ K. These parameter values also reproduce the measured temperature dependence of the frequency shift $\delta f(T) = f(T) - f(0)$ and quality factor $Q(T)$ of the resonance, according to $\delta f(T)/f(0) = -\alpha \delta L_s(T)/2L_s(0)$ and $Q^{-1}(T) = Q^{-1}(0) + \alpha R_s(T)/\omega L_s(0)$.

α is the fraction of the circuit inductance that is contributed by the surface inductance L_s . For our test device (Fig. 2), which has $\alpha \approx 0.04$, $V = 2,000 \mu\text{m}^3$ and $\Delta \approx 0.171$ meV (Fig. 3), we expect $|\delta f|/f_0 \approx 5 \times 10^{-5}$ when a single 5.9-keV X-ray photon is absorbed (Fig. 4). This frequency shift should be compared against the width of the resonance, which is $\Delta f/f_0 = Q^{-1} = 2 \times 10^{-5}$, as the quality factor of our test device is $Q = 52,500$ at low temperatures. Thus, we conclude that single 5.9-keV X-ray photons should easily be detectable, driving the resonator nearly off-resonance, and producing large phase shifts of the order of 180° . In fact, this is what is observed (Fig. 4). In this example, the fraction of Cooper pairs that are broken is quite small, of the order of 2×10^{-3} . This allows us to regard the derivative $\partial Z_s / \partial n_{qp}$ as being essentially constant.

We studied a device with a quarter-wavelength transmission line resonator (Fig. 2). This circuit is the transmission line equivalent of the lumped element resonator shown in Fig. 1b. A homodyne detection scheme (Fig. 3, right inset) is used to measure the transmission amplitude and phase at the resonance frequency as a function of time. We convert the time-dependent transmission measurements into a phase angle, $\theta(t)$, as indicated in Fig. 3 (left inset). Figure 3 shows the measured and calculated phase angle versus N_{qp} , the total number of quasiparticles in the centre strip of the coplanar waveguide (CPW) line, obtained by varying the

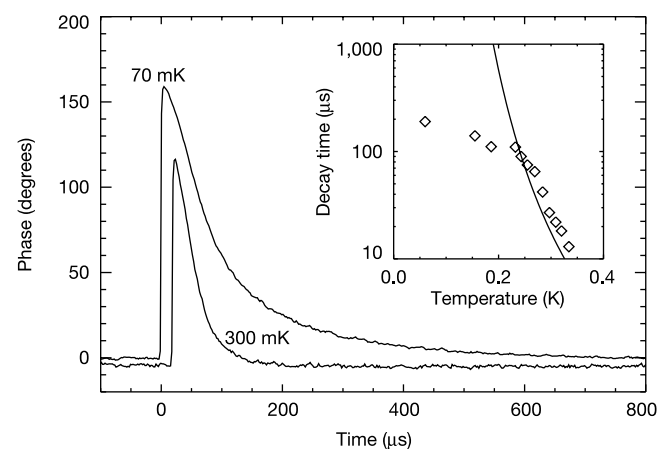


Figure 4 Single-photon X-ray pulses measured at 70 and 300 mK. The phase is calculated from the time-domain I/Q data (Fig. 3). The data points are taken at $2\text{-}\mu\text{s}$ intervals and are statistically independent. To compare pulse heights from absorption events in the centre strip to the thermal calibration of Fig. 3, we should multiply by $\alpha_{\text{centre}}/\alpha \approx 0.66$. On the other hand, the thermal quasiparticles are distributed uniformly throughout the device, whereas the largest X-ray events will be those from X-rays absorbed near the sensitive shorted end. These two effects approximately cancel. A simple exponential decay does not describe the pulses well, except in the tails. In the inset, we compare the decay times measured using the tails of the pulses to the theoretical quasiparticle recombination time (solid line), τ_{rec} , calculated by Kaplan²⁹. At high temperatures the measured decay times lie above the theoretical curve, but have a similar temperature dependence, whereas at low temperatures the measured decay times saturate. The measured decay times are expected to be larger than τ_{rec} owing to phonon trapping. The saturation could have a number of explanations: for example, an excess background quasiparticle population due to stray radiation; or trapping and recombination at defects, grain boundaries, or fluxons (due to Earth's field), where the gap parameter is reduced. Quasiparticles may also diffuse out of the centre strip through the shorted end of the resonator. The average diffusion length before recombination can be expected to be of the order of 1 mm. If the decay time saturation is due to excess background quasiparticles, we estimate $N_{qp} \approx 2 \times 10^6$ by calculating the thermal density at the saturation temperature. This value of N_{qp} would not significantly modify the coupling-limited low-temperature Q of the resonator, but is approximately consistent with an estimate of the internal resonator loss obtained using the depth of the transmission dip.

temperature. The responsivity $\partial\theta/N_{qp}$ is fairly constant for low quasiparticle concentrations, when Q is close to its low-temperature limit $Q(0)$. At higher quasiparticle densities, the phase responsivity decreases as the resonance moves off the excitation frequency and becomes broader.

We investigated the photoresponse of the device using an ^{55}Fe source, which emits 5.9-keV X-rays. Pulses of various amplitudes are observed, as expected, as the resonator has a position-dependent response owing to the standing-wave current distribution. Quasiparticle outdiffusion also plays a role (Fig. 2). The largest events are identified with hits on the centre strip near the shorted end. Sample X-ray events of this type measured at 70 mK and 300 mK are shown in Fig. 4. The X-rays create approximately $^{19}\eta E_{xray}/\Delta \approx 2 \times 10^7$ quasiparticles, which should saturate the detector (Fig. 3) and produce a peak phase shift approaching 180° , as observed.

We can estimate the noise equivalent power (NEP) of the detector as a function of frequency ω using the measured phase noise power spectrum $S_\theta(\omega)$ along with the measured responsivity $\partial\theta/\partial N_{qp}$ (Fig. 3):

$$\text{NEP}^2(\omega) = S_\theta(\omega) \left(\frac{\eta\tau_0}{\Delta} \frac{\partial\theta}{\partial N_{qp}} \right)^{-2} (1 + \omega^2\tau_0^2) \quad (1)$$

where τ_0 , the response time of the detector, is determined from the pulse tails. Figure 5 shows the estimated NEP at 100 mK, and the

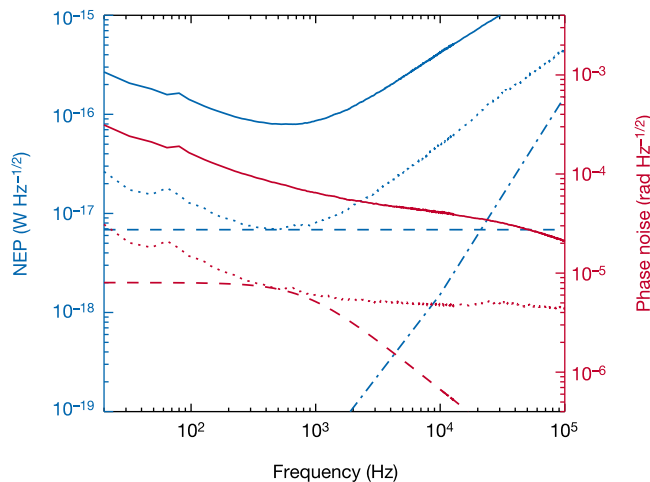


Figure 5 A plot of the noise measured for the test device; several other devices have given similar results. The right axis (red curves) gives the total output phase noise spectrum $S_\theta^{1/2}(\omega)$, while the left axis (blue curves) show noise equivalent power, NEP as described in the text. The solid lines give the total noise, and the dotted lines give the estimated contribution from the HEMT amplifier, measured by tuning the excitation frequency off resonance. The NEP values are applicable for absorbed power levels low enough that τ_0 and $\partial\theta/\partial N_{qp}$ are not changed. The dashed lines give an upper bound to the generation-recombination noise, assuming a quasiparticle population of $N_{qp} = 2 \times 10^6$ in the device, which is consistent with the measured saturation of pulse decay time (Fig. 4). The NEP contribution from the microwave synthesizer, estimated from its specifications, is shown as the dot-dashed line. At frequencies lower than $\Delta\omega_r$, the bandwidth of the resonator, synthesizer noise is suppressed by the factor $\omega/\Delta\omega_r$, because the resonator phase can track the synthesizer phase fluctuations within a bandwidth $\Delta\omega_r$. The estimated noise contributions appear to be at least 10 dB below the total measured noise, which indicates extra noise of unknown origin; the possibilities include random motion of fluxons trapped in the superconducting films, stray radiation with $h\nu \gg \Delta$, and substrate noise. These may be experimentally investigated by using magnetic shielding, improving the filtering for the microwave coaxial lines in the cryostat, and by varying the device geometry, materials, substrate, and so on. The amplifier noise should not be a limiting factor. Gain and phase fluctuations may be eliminated using carrier suppression or other methods. For the present device, the amplifier white noise (50 K) limit is $\Delta E = 1$ eV and $\text{NEP} = 5 \times 10^{-18} \text{ W Hz}^{-1/2}$, improving to 0.3 eV and $2 \times 10^{-18} \text{ W Hz}^{-1/2}$ for an optimized 5 K amplifier.

estimated contributions of the amplifier and synthesizer phase noise. The energy resolution can be estimated²¹ from $\text{NEP}(\omega)$, giving $\Delta E = 11$ eV, which is consistent with the pulse signal-to-noise ratio. It is important to note that the present device cannot actually be used for spectroscopy at this resolution owing to its position-dependent response; a more sophisticated device will be needed (see below). Nonetheless, this result indicates the potential of the concept.

The intrinsic noise at $T \ll T_c$ is expected to arise from random generation and recombination of quasiparticles^{22,23}, which should be negligible at 100 mK if the quasiparticle population is thermal. Although it is possible that there are excess quasiparticles—for example, generated by stray radiation—limits may be set on this excess population using the observed saturation of τ_0 (Fig. 4), which in turn gives a generation-recombination NEP limit of $6 \times 10^{-18} \text{ W Hz}^{-1/2}$ (Fig. 5). The measured noise appears to be larger than the estimated readout and intrinsic noise; this is clearly an important area for further research.

Frequency multiplexing is accomplished by coupling an array of many resonators with slightly different resonance frequencies to a common transmission line. A single cryogenic HEMT (high electron mobility transistor) amplifier is capable of amplifying the output signals from a large number of detectors, as many as 10^3 – 10^4 , depending on the amplifier bandwidth and the detector frequency spacing. This depends on the Q and the lithographic control of the resonances. An array of synthesizers and quadrature receivers at room temperature completes the system, and is readily implemented using miniature, low-cost, low-power integrated circuits developed for wireless communications.

Practical detectors will require efficient coupling of photons or quasiparticles into the sensitive end of the resonators. Fortunately, methods developed for other types of superconducting detectors may be reused. Planar antennas with microstrip lines²⁴ work well at millimetre wavelengths, and tantalum superconducting films provide excellent optical/ultraviolet/X-ray quantum efficiency^{4,8,14}. Diffusion and quasiparticle trapping^{14,25} allow the separation of absorber and detector functions, which provides considerable flexibility for detector optimization.

Although the demonstrated performance is already good enough for some applications—such as ground-based submillimetre imaging—better sensitivity is desirable. What are the fundamental limits? The amplifier noise contribution can be substantially reduced (Fig. 5), to $\text{NEP} \approx 10^{-18} \text{ W Hz}^{-1/2}$ for the present device. Further reductions of the readout noise will require a larger responsivity $\partial\theta/N_{qp}$, which is achievable by increasing Q or decreasing the volume; we have already measured $Q = 2 \times 10^6$ for more weakly coupled devices. Similarly, the intrinsic generation-recombination noise does not set a hard sensitivity limit, as it decreases exponentially with temperature²³. Thus, the potential exists to develop detectors with NEP below $10^{-19} \text{ W Hz}^{-1/2}$, which should be sufficient for most applications. The key issues for the future are understanding and reducing the noise sources in these devices, which we suspect are non-fundamental, and demonstrating efficient radiation coupling and multiplexing. □

Received 10 March; accepted 2 September 2003; doi:10.1038/nature02037.

1. Newbury, D. *et al.* The approaching revolution in X-ray microanalysis: The microcalorimeter energy dispersive spectrometer. *J. Radioanal. Nucl. Chem.* **244**, 627–635 (2000).
2. Booth, N., Cabrera, B. & Fiorini, E. Low-temperature particle detectors. *Annu. Rev. Nucl. Part. Sci.* **46**, 471–532 (1996).
3. Saab, T. *et al.* Deployment of the first CDMS II ZIP detectors at the Stanford Underground Facility. *Nucl. Phys. B* **110**, 100–102 (2002).
4. Peacock, A. *et al.* Single optical photon detection with a superconducting tunnel junction. *Nature* **381**, 135–137 (1996).
5. Cabrera, B. *et al.* Detection of single infrared, optical, and ultraviolet photons using superconducting transition edge sensors. *Appl. Phys. Lett.* **73**, 735–737 (1998).
6. Lee, S., Gildemeister, J., Holmes, W., Lee, A. & Richards, P. Voltage-biased superconducting transition-edge bolometer with strong electrothermal feedback operated at 370 mK. *Appl. Opt.* **37**, 3391–3397 (1998).
7. de Korte, P. Cryogenic imaging spectrometers for X-ray astronomy. *Nucl. Instrum. Meth. A* **444**(1–2),

- 163–169 (2000).
8. Rando, N. *et al.* S-cam: A spectrophotometer for optical astronomy: Performance and latest results. *Rev. Sci. Instrum.* **71**(12), 4582–4591 (2000).
9. Twerenbold, D. Giaever-type superconducting tunneling junctions as high-resolution X-ray-detectors. *Europhys. Lett.* **1**(5), 209–214 (1986).
10. Kraus, H. *et al.* High-resolution X-ray-detection with superconducting tunnel-junctions. *Europhys. Lett.* **1**(4), 161–166 (1986).
11. Irwin, K., Hilton, G., Wollman, D. & Martinis, J. X-ray detection using a superconducting transition-edge sensor microcalorimeter with electrothermal feedback. *Appl. Phys. Lett.* **69**(13), 1945–1947 (1996).
12. Irwin, K. *et al.* A Mo-Cu superconducting transition-edge microcalorimeter with 4.5 eV energy resolution at 6 keV. *Nucl. Instrum. Meth. A* **444**, 184–187 (2000).
13. Angloher, G. *et al.* Energy resolution of 12 eV at 5.9 keV from Al superconducting tunnel junction detectors. *J. Appl. Phys.* **89**, 1425–1429 (2001).
14. Li, L. *et al.* Improved energy resolution of X-ray single photon imaging spectrometers using superconducting tunnel junctions. *J. Appl. Phys.* **90**, 3645–3647 (2001).
15. Chervenak, J. *et al.* Superconducting multiplexer for arrays of transition edge sensors. *Appl. Phys. Lett.* **74**, 4043–4045 (1999).
16. Yoon, J. *et al.* Single superconducting quantum interference device multiplexer for arrays of low-temperature sensors. *Appl. Phys. Lett.* **78**, 371–373 (2001).
17. Mazin, B. A., Day, P. K., Leduc, H. G., Yavonakis, A. & Zmuidzinas, J. Superconducting kinetic inductance photon detectors. *Proc. SPIE* **4849**, 283–293 (2002).
18. Tinkham, M. *Introduction to Superconductivity*, 2nd edn (McGraw-Hill, New York, 1996).
19. Kozorezov, A. G. *et al.* Quasiparticle-phonon downconversion in nonequilibrium superconductors. *Phys. Rev. B* **61** (May), 11807–11819 (2000).
20. Mattis, D. C. & Bardeen, J. Theory of the anomalous skin effect in normal and superconducting metals. *Phys. Rev.* **111**, 412–417 (1958).
21. Moseley, S., Mather, J. & McCammon, D. Thermal detectors as x-ray spectrometers. *J. Appl. Phys.* **56**, 1257–1262 (1984).
22. Wilson, C., Frunzio, L. & Prober, D. Time-resolved measurements of thermodynamic fluctuations of the particle number in a nondegenerate Fermi gas. *Phys. Rev. Lett.* **87**, 067004 (2001).
23. Sergeev, A., Mitin, V. & Karasik, B. Ultrasensitive hot-electron kinetic-inductance detectors operating well below the superconducting transition. *Appl. Phys. Lett.* **80**, 817–819 (2002).
24. Zmuidzinas, J. & LeDuc, H. G. Quasi-optical slot antenna SIS mixers. *IEEE Trans. Microwave Theory Tech.* **MTT-40**, 1797–1804 (1992).
25. Li, L., Frunzio, L., Wilson, C. & Prober, D. Quasiparticle nonequilibrium dynamics in a superconducting Ta film. *J. Appl. Phys.* **93**, 1137–1141 (2003).
26. Chang, W. Inductance of a superconducting strip transmission-line. *J. Appl. Phys.* **50**, 8129–8134 (1979).
27. McMillan, W. Transition temperature of strong-coupled superconductors. *Phys. Rev.* **167**, 331–334 (1968).
28. Wells, G. L., Jackson, J. E. & Mitchell, E. N. Superconducting tunneling in single-crystal and polycrystal films of aluminum. *Phys. Rev. B* **1**, 3636–3644 (1970).
29. Kaplan, S. *et al.* Quasiparticle and phonon lifetimes in superconductors. *Phys. Rev. B* **14**, 4854–4873 (1976).

Acknowledgements This work has been supported in part by NASA (Aerospace Technology Enterprise), the JPL Director's Research and Development Fund, and the Caltech President's Fund. We are grateful for the support of A. Lidow, Caltech Trustee.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.K.D. (Peter.K.Day@jpl.nasa.gov).

Imaging coexisting fluid domains in biomembrane models coupling curvature and line tension

Tobias Baumgart¹, Samuel T. Hess² & Watt W. Webb¹

¹Applied and Engineering Physics, Cornell University, Ithaca, New York 14853, USA

²Laboratory of Cellular and Molecular Biophysics, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland 20892, USA

Lipid bilayer membranes—ubiquitous in biological systems and closely associated with cell function—exhibit rich shape-transition behaviour, including bud formation¹ and vesicle fission². Membranes formed from multiple lipid components can laterally separate into coexisting liquid phases, or domains, with distinct compositions. This process, which may resemble raft

formation in cell membranes, has been directly observed in giant unilamellar vesicles^{3,4}. Detailed theoretical frameworks^{5–11} link the elasticity of domains and their boundary properties to the shape adopted by membranes and the formation of particular domain patterns, but it has been difficult to experimentally probe and validate these theories. Here we show that high-resolution fluorescence imaging using two dyes preferentially labelling different fluid phases directly provides a correlation between domain composition and local membrane curvature. Using freely suspended membranes of giant unilamellar vesicles, we are able to optically resolve curvature and line tension interactions of circular, stripe and ring domains. We observe long-range domain ordering in the form of locally parallel stripes and hexagonal arrays of circular domains, curvature-dependent domain sorting, and membrane fission into separate vesicles at domain boundaries. By analysing our observations using available membrane theory, we are able to provide experimental estimates of boundary tension between fluid bilayer domains.

We study giant unilamellar vesicles (GUVs) formed from a ternary mixture of the lipids sphingomyelin, dioleoylphosphatidylcholine (DOPC) and cholesterol³. Sphingomyelin and cholesterol enrich in a liquid phase with short-range order (L_o), and DOPC prefers a disordered liquid (L_d) phase. The phase diagram of this lipid mixture at physiologically relevant temperatures shows a large binary coexistence region of L_o and L_d domains (A.K. Smith and G.W. Feigenson, personal communication; see also Supplementary Information for further details). Figure 1 presents images of equatorial sections of GUVs (obtained using two-photon microscopy), exhibiting phase coexistence of the L_o (blue) + L_d (red) phases at varying compositions. Figure 1a–d shows shapes with symmetry around an axis in the vertical direction within the image plane.

The geometry of phase-separated vesicles is theoretically obtained by minimizing an energy functional with contributions arising from bending resistance, lateral tensions, line tension and normal pressure difference. The bending energy F_b of an axially symmetric lipid membrane has components arising from mean curvature (first term) and Gauss curvature^{12,13} (second term), and is summed over every domain i of the vesicle⁸:

$$F_b^i = \frac{\kappa^i}{2} \int_{A^i} (C_m + C_p - C_0)^2 dA + \kappa_G^i \int_{A^i} C_m C_p dA \quad (1)$$

Here κ , κ_G , C_m and C_p are mean and Gauss bending rigidities and principal curvatures along the meridians and parallels, respectively, and the integral is extended over all domain areas i . C_0 is a curvature preference, which is often assumed to be constant within a domain^{5,8,11} but can in general vary locally¹⁴. The bending modulus κ typically has a value of $\sim 10^{-19}$ J (ref. 15). The second term in equation (1) influences the vesicle shape only if κ_G^i values differ between domains i (ref. 8).

Comparison of Fig. 1a and 1b, which show similar shapes even though the domains are reversed, suggests that bending resistance differences (or any curvature preferences) are not dominant in determining global vesicle shapes. In fact, a line tension σ at the boundaries between coexisting fluid membrane domains has been proposed to control membrane deformation, budding and fission⁵. Theoretical estimates of σ are of the order of 10^{-11} N for systems far from⁵ critical points of the lipid phase diagram, and σ vanishes at critical points^{5,16}. Above a limiting boundary length $s^0 \approx 8\kappa/\sigma \approx 80$ nm, an initially flat domain would transform into a complete spherical bud with vanishing neck radius, provided that sufficient membrane area is available⁵. Accordingly, the domain boundary radii observed in Fig. 1 favour bud formation even for much smaller estimates of σ .

This simplifying theoretical description neglects the constraints on membrane domain areas and surface-to-volume ratio (s/v), which are constant on the timescale of our experiment (up to

1 h), and also neglects the membrane geometry in the neck region of a budded domain. We show (see Supplementary Information) that for liposomes such as shown in Fig. 1, line tension drives vesicles into a shape close to a limit shape characterized by a minimum domain boundary radius R_b . This limit shape consists of truncated spheres, connected at the phase boundary, with areas and volumes of the spherical caps determined by the constrained vesicle volume and domain areas. A further transition into a shape with vanishing neck radius (with vanishing mean curvature and line energy) is suppressed by a lateral membrane domain tension Σ^i . According to Fig. 1a, b and g, the meridional radius of curvature in the neck

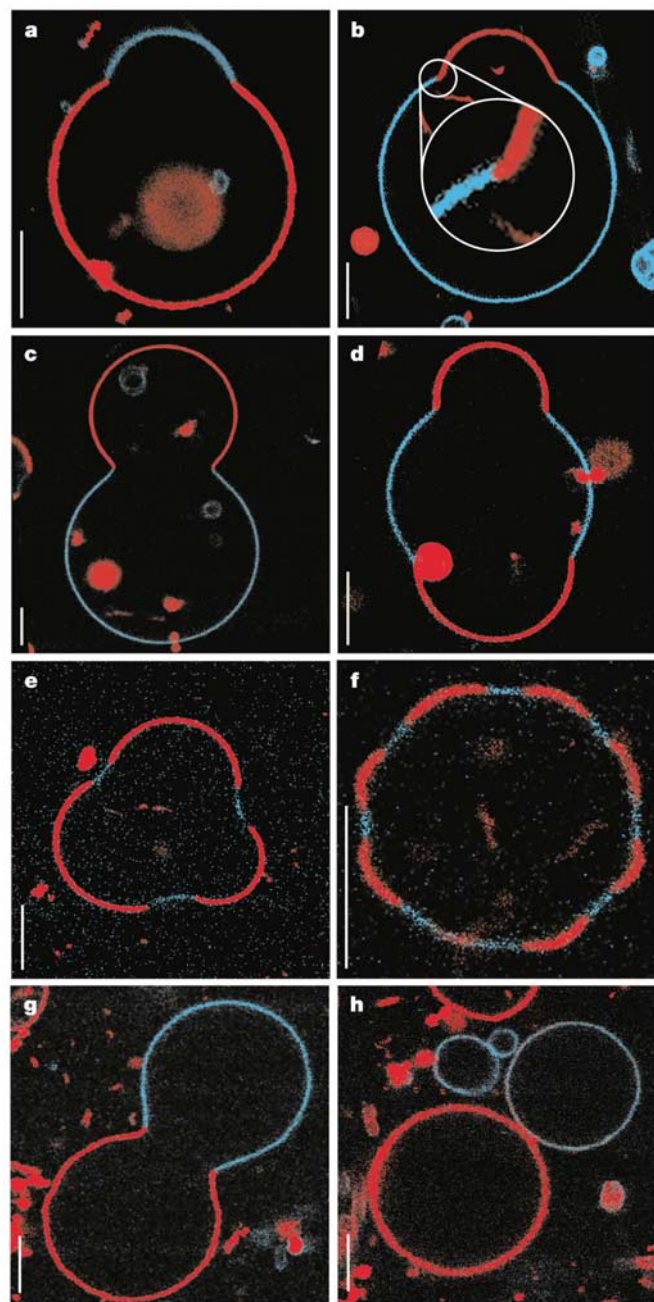


Figure 1 Two-photon microscopy images showing equatorial sections of GUVs with $L_o + L_d$ phase coexistence. GUVs were labelled with perylene (blue) in L_o phases and rho-DPPE (red) in L_d phases. All images superimpose red and blue channels (for separations, see Supplementary Fig. 1'). Some GUVs contain smaller vesicles and lipid fragments. Images obtained at 25 °C, except for **g** (30 °C) and **h** (35 °C). Inset of **b**, enlarged neck region. Scale bars, 5 μ m.

region is close to the limit of optical resolution. Many of the vesicles we analysed, however, suggest a high, but finite reverse meridional curvature in this region (see, for example, Fig. 1b inset) with a continuous membrane geometry (as opposed to a kink), a reasonable assumption for domain boundaries of finite thickness within fluid membranes. In the neck region, L_d domains bend towards the L_o domains (as shown, for example, in Fig. 1b and g), suggesting smaller bending rigidities κ of L_d compared to L_o phases¹⁵. The smaller rigidity (or higher flexibility) of L_d phases agrees with our observation that the L_o phase exhibits smaller amplitudes of thermally excited out-of-plane undulations^{15,17}.

A refined shape theory of Jülicher and Lipowsky⁸ allowed us to devise a fitting routine using five free parameters, and to determine mechanical vesicle parameters (see Supplementary Information for details). To establish the fundamental relation between domain composition and mechanical properties requires the currently undetermined tie lines of domain coexistence in the ternary phase diagram of our lipid mixtures. We therefore focus, as an example, on the shape of the vesicle shown in Fig. 1c. There we obtain a line tension $\sigma \approx 9 \pm 0.3 \times 10^{-13}$ N, which is an order of magnitude smaller than Lipowsky's above-mentioned rough estimate, which is based on an estimation of the energetic penalty arising from interfacial free energies of a cut across the lipid bilayer at the phase boundary⁵. However, our experimentally obtained value lies in the range of another theoretical estimate (P. Kuzmin, S. Akimov, F. Cohen and J. Zimmerberg, personal communication), which minimizes the sum of energies from elastic compression/stretching and tilt, and hydrophobic height mismatch between the L_o and L_d membrane phases.

Lateral tensions obtained were $\Sigma^L_d \approx 8.2 \pm 0.1 \times 10^{-5}$ mN m⁻¹ and $\Sigma^L_o \approx 1.06 \pm 0.01 \times 10^{-4}$ mN m⁻¹, balanced by a pressure difference (outer minus inner pressure) of $P \approx -1.83 \pm 0.01 \times 10^{-2}$ N m⁻². The lateral tensions significantly suppress microscopically visible thermally excited out-of-plane undulations¹⁸, consistent with our observations. We obtain $\kappa^{L_o}/\kappa^{L_d} \approx 1.25 \pm 0.60$, again suggesting a slightly higher bending rigidity of L_o phases. In contrast to the above-mentioned fitting parameters, the value of $\kappa^{L_o}/\kappa^{L_d}$ is, however, associated with a high uncertainty. Our fit shows a small systematic deviation, particularly in the neck region, which is readily understood by (1) the uncertainty of the experimental determination of the neck geometry, which is close to the optical resolution, and (2) the simplifying assumptions of the theory⁸, that is, a constant membrane composition throughout a domain. In the case of the presence of cone-shaped lipids (like cholesterol), steep local curvature gradients can cause lateral and interleaflet lipid redistribution^{14,19} and therefore locally modify the mechanical membrane properties.

If the s/v ratio is further increased, either by changing osmotic stress or by a temperature increase, we observe the formation of complete spherical buds, that is, closure of the neck at the domain edge²⁰, and the fission into separate vesicles (Fig. 1g, h shows a fission sequence). This occasional shape transition occurs abruptly (timescale less than 0.5 s; see Supplementary Movie), indicating an energy barrier^{5,8}. In the case of neck radii R_b approaching molecular dimensions, our experimental estimate of σ indicates that the lateral tension $\Sigma = \sigma/R_b$ will rise to values where membranes are expected to rupture⁵. Indeed, we observe fission at the phase boundary, leaving vesicles that differ in their phase state and relative lipid composition (see Fig. 1h and Supplementary Movie).

In addition to phase-separated vesicles with one spheroidal domain, we find vesicles with two, three and multiple caps (Fig. 1d–f), often coexisting within the same sample (vesicles shown in Fig. 1c–f were prepared from a lipid mixture of the same composition). It is possible that those liposomes with multiple caps lacking a curvature preference are not in a state of global energy minimum, as we frequently observe the dynamics of fusion of domains^{4,11,21}, which reduces total line energy. In vesicles with

budding domains, however, this fusion step seems to be impeded, because domains tend to repel each other upon approach, avoiding high curvature build-up. Indeed, equatorial section images through GUVs with multiple caps (Fig. 1d–f; in these images the L_o phase is continuous and contains separated spheroidal L_d phase caps), as well as stacks of focal plane images of vesicle hemispheres showing domains of similar size (Fig. 2a and b) reveal approximate long-range ordering patterns of cap-shaped domains. Figure 2a and b shows arrangements into both hexagonal and inverted hexagonal patterns, depending on membrane composition. Long-range order-

ing of fluid membrane domains is clearly matched by a curvature pattern (Figs 1d–f, and 2a and b).

Domains imaged at room temperature are circular (Fig. 2a and b)^{3,4,21}. When temperature was increased to values several degrees below the mixing/demixing transition temperature T_m , in vesicles with and without excess curvature, we observed both L_o phase and L_d phase circular domains to undulate laterally, indicating reduced line tension at higher temperatures. Further temperature increase led to the formation of a homogeneous membrane. Thermally excited lateral domain undulations have normal mean square amplitudes $\langle u_n^2 \rangle = k_B T / (\pi r_o \sigma (n^2 - 1))$, where k_B is Boltzmann's constant, T is temperature, r_o is the average domain radius and n is the mode number. Accordingly, domains with a radius $r_o \approx 5 \mu\text{m}$, which visibly undulate with optically resolvable amplitudes, indicate a line tension below $\sigma \approx 10^{-14} \text{ N}$ (at high temperatures), which is two orders of magnitude smaller than the room-temperature value determined from our shape fitting. In-plane fluctuations of domains at temperatures close to, but below, T_m tend to increase when approaching a composition range (mole fractions of sphingomyelin/DOPC/cholesterol) of $0.63 \pm 0.035 / 0.07 \pm 0.035 / 0.3 \pm 0.05$. From preliminary data on the temperature dependence of $L_o + L_d$ phase coexistence (see Supplementary Information for further details), we assume this region to be near an upper critical mixing/demixing point, where line tension is expected to decrease as $\sigma \propto (T_c - T)^\lambda$ (here T_c is the critical temperature, and the critical exponent $\lambda = 1$; refs 5, 16).

Within this composition range, where L_d phase domains in a continuous L_o matrix are found, at temperatures less than one degree below T_m , in vesicles with excess area (compared to a sphere), the undulating domains tend to assume strongly prolate elliptical shapes; upon further increasing T , a shape instability leads to a stripe-out⁷, that is, the formation of numerous thin stripes, with varying (fluctuating) lengths, which rapidly undulate in the plane of the membrane, and can span the whole liposome. We observed multiple stripe domains to arrange into laterally undulating patterns of locally parallel stripes. Close to T_m , scanning microscopy was difficult to apply to image these fluctuating patterns. In favourable cases, however, rapid cooling by several degrees led to a sufficient suppression of stripe pattern undulations (and a slight thickness increase of stripes) that the patterns could be imaged (Fig. 2c). Again, an equatorial section through the same GUV (Fig. 2d) shows a pattern of membrane curvature matching the phase pattern. Two different pattern defects are observed (Fig. 2c), namely the formation of loops and bifurcations with an angle near 120° . The relative thickness of stripes depends on membrane composition and temperature (Fig. 2e).

The stripe-out that we describe here resembles the rippling instability of collapsing bubbles²², but differs from stripe phase formation in polar lipid monolayers at the air/water interface²³, as the long-range dipolar interaction of zwitterionic lipids is significantly screened in the aqueous environment. Both an increase in total membrane area and the reduced line tension around buds near the main transition temperature will reduce lateral tension within the membrane, and favour the instability described above. Equilibrium stripe phases associated with membrane deformation were also predicted^{6,9,10}. Further theoretical study and systematic experiments are clearly required to fully understand the physics of this complex phenomenon. Fusion of the ends of a stripe is frequently observed, which upon cooling (and thereby increasing the line tension) leads to (possibly metastable) liposome-spanning ring domains (Fig. 2e and f), where the line tension dominates the shape of these vesicles.

Homogenous membranes with periodic curvature modulation, such as pearling states in tubular membranes²⁴ and 'starfish' vesicles²⁵ at high s/v ratios, have previously been reported. We frequently observed these types of homogeneous membranes at temperatures above T_m . Cooling slightly below T_m leads to the

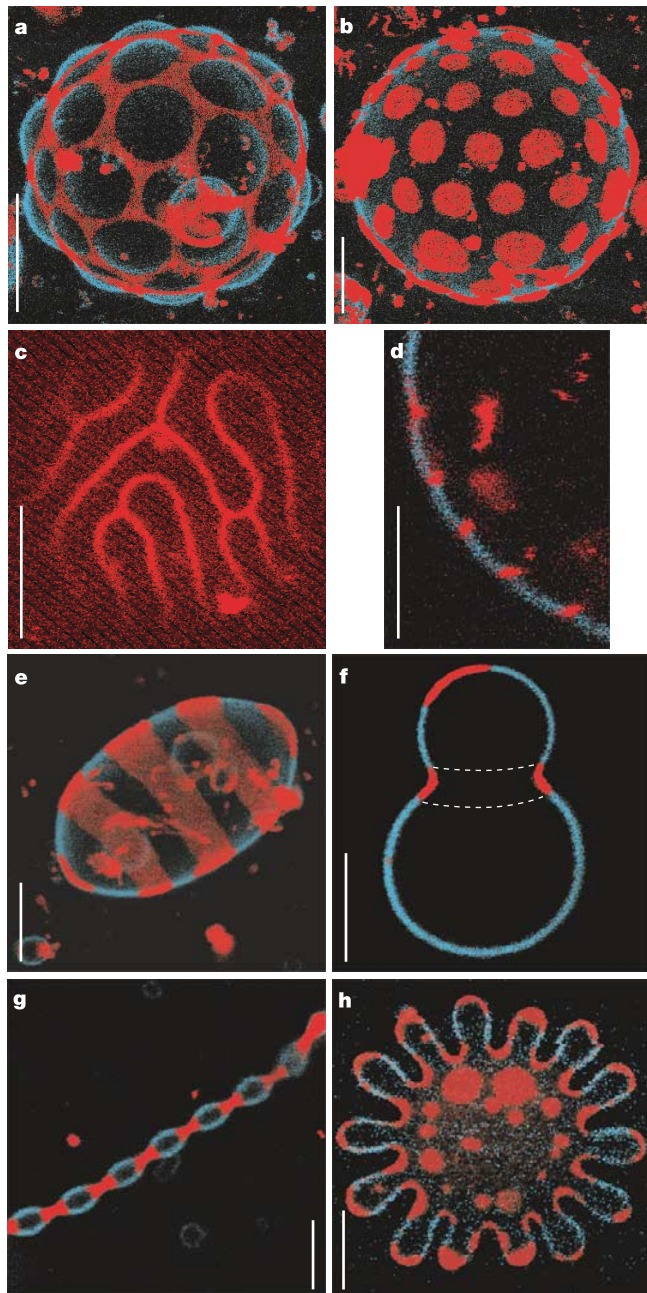


Figure 2 Two-photon microscopy images of GUVs with $L_o + L_d$ phase coexistence. All are superpositions of red and blue channels (for separations see Supplementary Fig. 2'), except **c** (red channel only). **d**, Equatorial section of the same vesicle as in **c**.

a, b, e, Hemispherical projections of image stacks taken at $0.5 \mu\text{m}$ spacing. Image in **c** is focused on vesicle top; **d, f, g** and **h** show equatorial sections. Dashed lines in **f** indicate an axially symmetric ring domain. Images are obtained at: **a** and **b**, 25°C ; **c** and **d**, 40°C ; **e**, 30°C ; **f**, 25°C ; **g**, 44°C ; **h**, 50°C . Scale bars, $5 \mu\text{m}$.

formation of numerous small domains, which accumulate and fuse into recurrent patterns, suggesting a curvature preference of coexisting domains (Fig. 2g and h). In both cases, the L_d phase is observed to favour saddle shapes, and the high-curvature tip region of the arms of the starfish vesicle (Fig. 2h), which again indicates smaller magnitudes of bending rigidities of L_d phases compared to L_o phase domains, which preferentially segregate into the lower-curvature tubular areas (Fig. 2g and h). Owing to the high temperature and the particular membrane composition, line tension effects are expected to be small in both cases. Note that domains in the essentially flat central body of the starfish vesicle are randomly distributed, as they do not experience a significant curvature gradient.

The experimental approach and observations described here allow us to address two-dimensional critical phenomena, the influence of membrane additives on line tension, and to test and advance theories on membrane shape. Some of our observations may furthermore have analogues in fundamental biological membrane processes^{26,27}. □

Methods

Vesicle preparation

GUVs were mostly prepared by the method of electrosweeling²⁸, at a temperature of 60 °C, in a solution of 100 mM sucrose. Using GUVs prepared in 50 mM KCl led to virtually the same phenomena observed at zero ionic strength. Lipid mixtures composed of varying fractions of sphingomyelin (from egg), dioleoylphosphatidylcholine and cholesterol were used. The elevated temperature ensured that GUVs were formed from swollen lipid membranes above the upper critical mixing/demixing temperature of L_o/L_d phase coexistence. The mixing/demixing temperature, T_m , varied with composition (see Supplementary Information for further details). Phase coexistence was observed at temperatures up to 56 ± 0.2 °C.

Fluorescence labelling

GUVs were labelled with two differentially partitioning membrane probes in order to distinguish between L_o and L_d domains. The headgroup labelled lipid probe N-lissamine rhodamine dipalmitoylphosphatidylethanolamine (rho-DPPE) partitioned so effectively into L_d domains that in many cases (depending on the membrane composition) no fluorescence in the coexisting L_o phase domains could be detected. The same behaviour was found for numerous other fluorescence-labelled lipid or fatty acid analogues, including the widely used amphiphilic indocarbocyanine dye DiI. We found that the polycyclic aromatic hydrocarbon dye perylene (among other polycyclic aromatic dyes of similar type) partitions preferentially into the L_o phase. The partitioning is, however, weaker compared to rho-DPPE. Rho-DPPE was added at 1/1,000 (dye/lipid). Perylene was added at 1/500. The only charged component in the lipid mixture used in the present study is the fluorescence probe rho-DPPE. In independent experiments using an uncharged lipid label, we confirmed the occurrence of modulated phases.

Two-photon fluorescence microscopy

Two-photon microscopy was performed as described²⁹, at $\lambda = 750$ nm, using a 60× water immersion objective. The temperature of GUVs on the microscope stage was controlled by means of a small water bath attached to the objective of an inverted microscope. The microscope objective was additionally kept at constant temperature and thermally isolated from the microscope stage. The temperature variation of our set-up is less than ± 100 mK.

Vesicle compositions

Lipid compositions are expressed as mole fractions of sphingomyelin, DOPC and cholesterol, respectively. Vesicles shown in Fig. 1: 1a, 0.25/0.5/0.25; 1b, 0.585/0.1/0.315; 1c–f, 0.45/0.45/0.1; 1g and h, 0.615/0.135/0.25. Vesicles shown in Fig. 2: 2a, 0.56/0.24/0.2; 2b, 0.615/0.135/0.25; 2c and d, 0.675/0.075/0.25; 2e, 0.638/0.112/0.25; 2f, 0.675/0.075/0.25; 2g, 0.63/0.07/0.3; 2h, 0.585/0.102/0.313.

Received 25 March; accepted 19 August 2003; doi:10.1038/nature02013.

- Käs, J. & Sackmann, E. Shape transitions and shape stability of giant phospholipid vesicles in pure water induced by area-to-volume changes. *Biophys. J.* **60**, 825–844 (1991).
- Döbereiner, H. G., Käs, J., Noppl, D., Sprenger, I. & Sackmann, E. Budding and fission of vesicles. *Biophys. J.* **65**, 1396–1403 (1993).
- Dietrich, C. *et al.* Lipid rafts reconstituted in model membranes. *Biophys. J.* **80**, 1417–1428 (2001).
- Veatch, S. L. & Keller, S. L. Organization in lipid membranes containing cholesterol. *Phys. Rev. Lett.* **89**, 268101 (2002).
- Lipowsky, R. Budding of membranes induced by intramembrane domains. *J. Phys. II France* **2**, 1825–1840 (1992).
- Leibler, S. & Andelman, D. Ordered and curved meso-structures in membranes and amphiphilic films. *J. Phys.* **48**, 2013–2018 (1987).
- Seul, M. & Andelman, D. Domain shapes and patterns: The phenomenology of modulated phases. *Science* **267**, 476–483 (1995).
- Jülicher, F. & Lipowsky, R. Shape transformations of vesicles with intramembrane domains. *Phys. Rev. E* **53**, 2670–2683 (1996).

- Andelman, D., Kawakatsu, T. & Kawasaki, K. Equilibrium shape of two-component unilamellar membranes and vesicles. *Europhys. Lett.* **19**, 57–62 (1992).
- Jiang, Y., Lookman, T. & Saxena, A. Phase separation and shape deformation of two-phase membranes. *Phys. Rev. E* **61**, R57–R60 (2000).
- Kumar, P. B. S., Gompper, G. & Lipowsky, R. Budding dynamics of multicomponent membranes. *Phys. Rev. Lett.* **86**, 3911–3914 (2001).
- Helfrich, W. Elastic properties of lipid bilayers: Theory and possible experiments. *Z. Naturforsch.* **28c**, 693–703 (1973).
- Jenkins, J. T. Static equilibrium configurations of a model red blood cell. *J. Math. Biol.* **4**, 149–169 (1976).
- Seifert, U. Curvature-induced lateral phase separation in two-component vesicles. *Phys. Rev. Lett.* **70**, 1335–1338 (1993).
- Duwe, H. P. & Sackmann, E. Bending elasticity and thermal excitations of lipid bilayer vesicles: Modulation by solutes. *Physica A* **163**, 410–428 (1990).
- Benvegnu, D. J. & McConnell, H. M. Line tension between liquid domains in lipid monolayers. *J. Phys. Chem.* **96**, 6820–6824 (1992).
- Schneider, M. B., Jenkins, J. T. & Webb, W. W. Thermal fluctuations of large quasi-spherical bimolecular phospholipid vesicles. *J. Phys.* **45**, 1457–1472 (1984).
- Helfrich, W. & Servuss, R. M. Undulations, steric interactions and cohesion of fluid membranes. *Nuovo Cimento D* **3**, 137–151 (1984).
- Chen, C.-M., Higgs, P. G. & MacKintosh, F. C. Theory of fission for two-component lipid vesicles. *Phys. Rev. Lett.* **79**, 1579–1582 (1997).
- Lipowsky, R. & Dimova, R. Domains in membranes and vesicles. *J. Phys. Condens. Matter* **15**, S31–S45 (2003).
- Samsonov, A. V., Mihalov, I. & Cohen, F. S. Characterization of cholesterol-sphingomyelin domains and their dynamics in bilayer membranes. *Biophys. J.* **81**, 1486–1500 (2001).
- Debregeas, G., de Gennes, P.-G. & Brochard-Wyart, F. The life and death of “bare” viscous bubbles. *Science* **279**, 1704–1707 (1998).
- Andelman, D., Brochard, F. & Joanny, J. F. Phase transitions in Langmuir monolayers of polar molecules. *J. Chem. Phys.* **86**, 3673–3681 (1987).
- Bar-Ziv, R. & Moses, E. Instability and “pearling” states produced in tubular membranes by competition of curvature and tension. *Phys. Rev. Lett.* **73**, 1392–1395 (1994).
- Wintz, W., Döbereiner, H. G. & Seifert, U. Starfish vesicles. *Europhys. Lett.* **33**, 403–408 (1996).
- Mukherjee, S. & Maxfield, F. R. Role of membrane organization and membrane domains in endocytic lipid trafficking. *Traffic* **1**, 203–211 (2000).
- Huttner, W. B. & Zimmerberg, J. Implications of lipid microdomains for membrane curvature, budding and fission. *Curr. Opin. Cell Biol.* **13**, 478–484 (2001).
- Mathivet, L., Cribier, S. & Devaux, P. F. Shape change and physical properties of giant phospholipid vesicles prepared in the presence of an AC electric field. *Biophys. J.* **70**, 1112–1121 (1996).
- Denk, W., Strickler, J. H. & Webb, W. W. Two-photon laser scanning fluorescence microscopy. *Science* **248**, 73–76 (1990).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank G. W. Feigenson, J. T. Jenkins, G. Gompper, J. Zimmerberg, A. K. Smith and A. T. Hammond for discussions. This work was supported in part by a CMBSTD grant of the W.M. Keck Foundation and an NIBIB-NIH grant to the Developmental Resource for Biophysical Imaging Opto-Electronics.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to W.W.W. (www2@cornell.edu).

Cool Indonesian throughflow as a consequence of restricted surface layer flow

Arnold L. Gordon, R. Dwi Susanto & Kevin Vranes

Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York 10964, USA

Approximately $10 \text{ million m}^3 \text{ s}^{-1}$ of water flow from the Pacific Ocean into the Indian Ocean through the Indonesian seas¹. Within the Makassar Strait, the primary pathway of the flow², the Indonesian throughflow is far cooler than estimated earlier, as pointed out recently on the basis of ocean current and temperature measurements^{3,4}. Here we analyse ocean current and stratification data along with satellite-derived wind measurements, and find that during the boreal winter monsoon, the wind

drives buoyant, low-salinity Java Sea surface water into the southern Makassar Strait, creating a northward pressure gradient in the surface layer of the strait. This surface layer 'freshwater plug' inhibits the warm surface water from the Pacific Ocean from flowing southward into the Indian Ocean, leading to a cooler Indian Ocean sea surface⁵⁻⁷, which in turn may weaken the Asian monsoon⁸. The summer wind reversal eliminates the obstructing pressure gradient, by transferring more-saline Banda Sea surface water into the southern Makassar Strait. The coupling of the southeast Asian freshwater budget to the Pacific and Indian Ocean surface temperatures by the proposed mechanism may represent an important negative feedback within the climate system.

The Indonesian throughflow (ITF) modifies the heat and freshwater budgets and air-sea heat fluxes of the Pacific and Indian oceans, and may exercise a role in the El Niño/Southern Oscillation (ENSO) and Asian monsoon climate phenomenon⁵⁻⁹. Observations show that the water composing the ITF is derived from the North Pacific thermocline^{2,10}, though at greater depth (water cooler than 6 °C) it is directly drawn from the South Pacific^{2,10,11} (Fig. 1). In the past, the ITF temperature has been thought to be warm, 22 °C (ref. 12) to 24 °C (refs 13, 14). Recent measurements of the ITF within the Makassar Strait suggest that the transport-weighted

temperature is at least 9 °C colder than earlier estimates⁴.

The Indian-to-Pacific flow south of Tasmania, which balances the ITF, is cool, perhaps 6 to 8 °C (ref. 7). If this water were to flow directly into the Indonesian seas without atmospheric contact, the ITF is expected to be similarly cool. However, water destined for the ITF criss-crosses the tropical Pacific within the upwelling regimes of the zonal equatorial currents, to enter the North Pacific subtropics before ultimate export to the Indian Ocean as the ITF², offering ample opportunity for substantial warming by the atmosphere. For example, warming of a 10-Sv ($1 \text{ Sv} = 10^6 \text{ m}^3 \text{ s}^{-1}$) stream from 7 °C to a typical upper thermocline temperature of 23 °C over a period of 10 to 100 yr (a reasonable range of Pacific Ocean residence times for ITF-destined water) requires mere fractions of a W m^{-2} , which is insignificant relative to the low-latitude heat exchange between the ocean and atmosphere of tens of W m^{-2} . Thus it is likely that the cool stream from south of Tasmania would be substantially heated en route to the Indonesian seas. Inverse solutions¹⁵⁻¹⁷ assume a steady-state ocean while using oceanographic sections obtained at different seasons or years, and may not be able to detect the subtleties of circulation pathways and heating of ITF waters within the Pacific Ocean. Although the ITF coolness may be a direct consequence of the large-scale circulation, as shown below it is more probably the result of a more localized mechanism within the

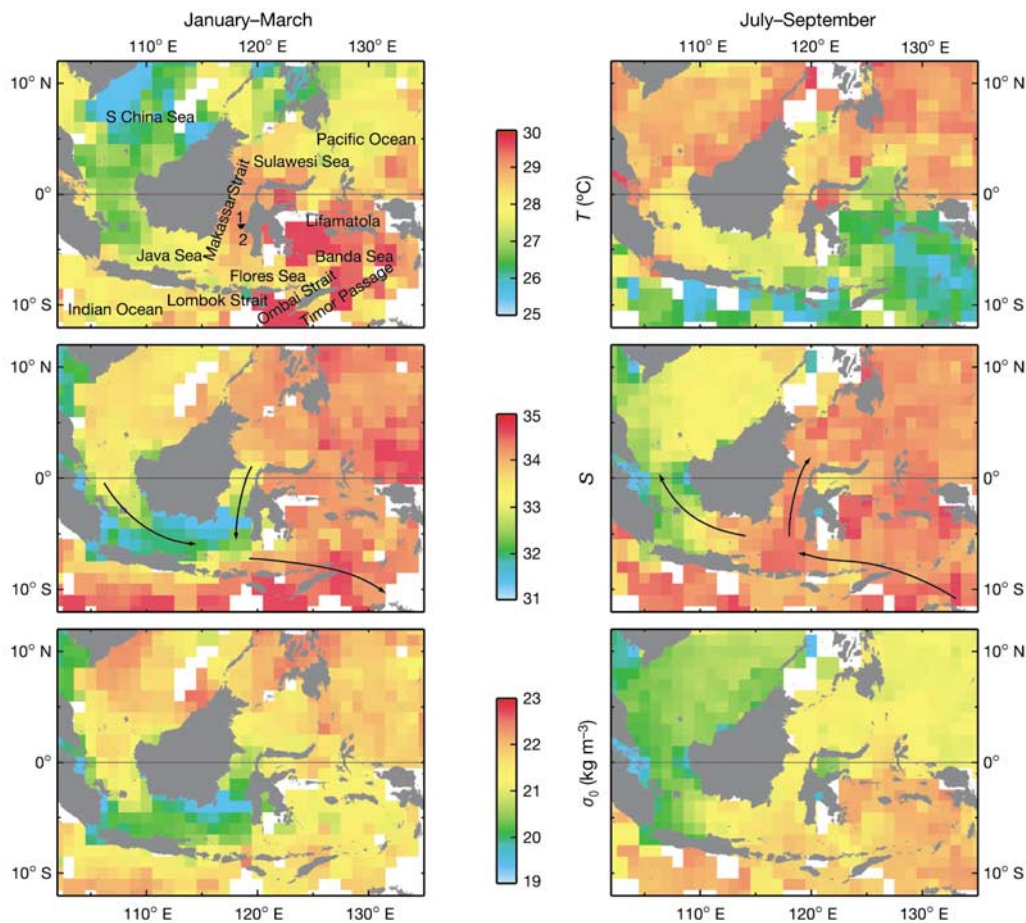


Figure 1 The Indonesian seas. Shown are the upper 20-m averages of summer and winter temperature (T), salinity (S) and density anomaly (σ_t) of the Indonesian seas (data from ref. 29). The prevailing monsoon winds are shown on the salinity panels; place names and the location of the two Makassar Strait moorings are shown on the January–March temperature panel. On 23 November 1996, the current meter mooring no. 1 was deployed at 2° 52' S, 118° 27' E in the 2,137 m deep, 45 km wide, Labani Channel of the Makassar Strait. Mooring no. 2 was deployed on 30 November just east of MAK-1 at 2° 51' S, 118° 38' E (ref. 18). The MAK moorings also included temperature sensors at 11

levels from 110 m and deeper³. During the northwest monsoon, January to March, the Java Sea low-salinity surface water shifts into the southern Makassar Strait. During the southeast monsoon, July to September, the southern Makassar Strait surface layer is more saline and less buoyant, as the southeast monsoon winds return the low-salinity water into the Java Sea. The buoyant surface water of the southern Makassar Strait inhibits southward transport within the surface layer during the northwest monsoon, despite southward wind within the Strait, lowering the temperature of the Indonesian throughflow.

Indonesian seas, which inhibits the transfer of warm Pacific surface water into the Indian Ocean.

Two moorings deployed in the Makassar Strait¹⁸ (Fig. 1) carried upward-looking acoustic Doppler current profilers (ADCPs) set at 150 m, with four current meters at 200 m, 250 m, 350 m and 750 m. Though the moorings measured the ITF for 1.7 yr, owing to instrument failure the combined ADCP time series measured the surface layer flow for 7 months, from December 1996 through to June 1997. MAK-2 ADCP time series spans 3 months, ending in early March, and the MAK-1 ADCP (whose data were recently processed, overcoming a data recording problem) spans 7 months, ending in early July.

The Makassar Strait transport for 1997 was 9.3 Sv (ref. 18). This compares favourably with export to the Indian Ocean within the upper 680 m (sill depth of the Makassar Strait) of 7.3–10.7 Sv through the passages of the Sunda Islands: 5.6–9.0 Sv for the combined Timor Passage and Ombai Strait^{19,20}, and the Lombok Strait transport of 1.7 Sv (ref. 21). Although the measurements cited were made at different times, an ITF of 10 Sv or slightly larger seems reasonable.

The interocean heat transfer through the Indonesian seas depends on the relationship of the volume transport profile to the temperature profile. Estimates based on direct measurements of currents and temperature in the Makassar Strait from the MAK moorings argue for an ITF transport-weighted temperature of slightly less than 15 °C (ref. 4). It is unlikely that the Makassar Strait transport-weighted temperature can be increased from 15 °C to 24 °C before export to the Indian Ocean via the Sunda passages, as such heating requires an atmosphere-to-ocean heat flux of

580 W m⁻², an order of magnitude greater than the estimated air-sea heat flux of 50 W m⁻² for this region²².

The relatively cool ITF results from a restricted contribution of warm surface waters (Fig. 2). Within the ADCP record, the surface layer transport varies strongly with season, with low southward speeds interspersed with periods of northward flow during the northwest monsoon from December 1996 to March 1997. From April to the end of the ADCP record in early July 1997, the surface flow is primarily southward. Even when southward, the surface layer transport is never stronger than the thermocline (100–200 m) transport. From April to June 1997, during strong southward surface layer flow, the average transport within each of the two 50-m slabs between 100 and 200 m is 1.23 Sv, compared to 1.09 Sv in the upper 50 m.

We propose that a major factor in restricting a surface layer contribution to the ITF is the meridional buoyancy gradient set up within the Makassar Strait that accompanies the redistribution of low salinity Java Sea water (Fig. 1; Fig. 3). The low surface salinities of the Java and South China seas are due to heavy precipitation and large river runoff from Southeast Asia. During the northwest monsoon, the eastward zonal wind expels the low salinity, buoyant water of the shallow Java Sea into the surface layer of the southern Makassar Strait. During the remainder of the year, the winds are either weak (monsoon transition months) or directed westward (southeast monsoon), drawing more-saline surface waters of the Flores and Banda seas into the Java Sea, shifting buoyancy from the southern Makassar Strait into the western Java Sea and South China Sea. The seasonal change of surface layer buoyancy within the Sulawesi Sea at the northern end of the Makassar Strait is much

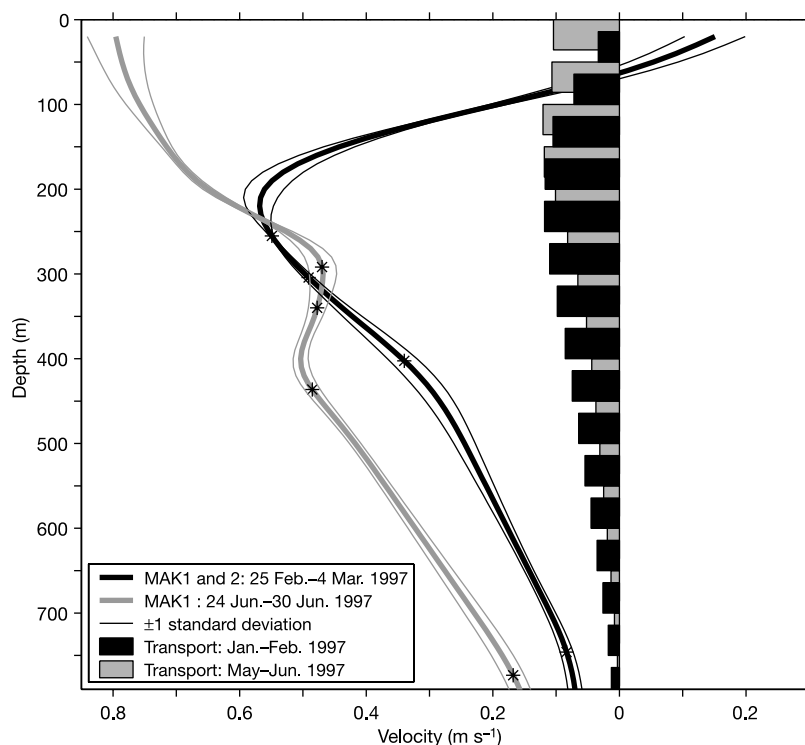


Figure 2 Profile of meridional velocity and transport as measured from mooring no. 1 (Fig. 1). The velocity profile (southward shown negative) with one standard deviation is shown for two periods: from 25 February to 4 March 1997 (the northwest monsoon, thick black line) and from 24 to 30 June 1997 (early southeast monsoon, thick grey line). The asterisks show the observed mean current and depth recorded at the four current meters for each of the two time periods. Above 150 m the curves are defined by the ADCP data. The southward transport in Sv (1 Sv = 10⁶ m³ s⁻¹) for 50 m depth bins is shown as a bar

graph for January and February 1997 (black bars) representing the northwest monsoon, and for May and June 1997 (grey bars) representing the early southeast monsoon. The reduced surface layer southward flow during the northwest monsoon is evident in both the velocity and transport profiles. The velocity profiles are not shown for depths less than 20 m, as the surface data bins are contaminated by sea surface scatter of the acoustical energy. For the top 50 m bin, transport is calculated by linear extrapolation to the surface.

more subdued (Fig. 3), as it is connected to the less seasonally variable surface salinity of the western Pacific. The more-buoyant surface water during the northwest monsoon generates a northward pressure gradient (relative to 100 decibars) within the Makassar Strait (Fig. 3), inhibiting southward flow within the surface layer. Warm Sulawesi Sea surface water that is inhibited from flowing southward within Makassar Strait during the northwest monsoon does not instead pass southward east of Sulawesi island, as the surface current charts for that region is northward flow in that season^{23,24}. Left unopposed, the meridional pressure gradient of the northwest monsoon would force surface water northward; however, the pressure gradient is opposed by the southward-directed wind (Fig. 1). The northward pressure gradient is removed during the southeast (summer) monsoon (Fig. 3). During the southeast monsoon the buoyancy-induced meridional pressure gradient is close to zero, but then the Makassar wind is directed towards the north (Fig. 1).

The observed Makassar Strait surface layer flow correlates with regional winds measured by the NSCAT satellite. The average 'relative' meridional current for the upper 50 m (Fig. 4) is directed towards the north from early December until late March, tracking the Java Sea zonal wind with a correlation of 0.8. The Java zonal wind becomes persistently eastward in early December 1996, and remains eastward until mid-March 1997. The Java wind leads the meridional surface currents by about ten days, probably the time it takes for a wind reversal in Java Sea to shift the low surface salinity water. Although the Java wind is clearly a dominant factor, the wind within the Makassar Strait does play a role in higher frequency variability of the surface flow. The Makassar surface layer flow displays a few strong northward events, notably in late December (event A in Fig. 4) and in mid-January (C in Fig. 4). In the first few

days of January there is a brief episode of strong southward flow (B in Fig. 4). The northward current events correlate at 3-day lag with weaknesses or reversals in the prevailing southward-directed winds. Similarly, the southward current period correlates with a strong southward wind event. The surface layer meridional current is weakly southward at the end of March, following the reversal of the Java Sea zonal winds by a week. The southward surface layer flow becomes predominant in mid-June as the westward Java Sea wind becomes more persistent.

The seasonal shift of low-salinity Java Sea water into the southern Makassar Strait acts as a 'freshwater plug' during the northwest monsoon, which effectively inhibits the Makassar Strait surface water from freely flowing southward. The cool ITF is a direct result of suppressed surface layer flow, as deeper, colder layers then dominate the full-column transport. During the southeast monsoon the 'freshwater plug' is removed, allowing southward transport in the surface layer, though northward wind is expected to hinder this effect.

Changes in the freshwater budget of the western Indonesian seas and Southeast Asian monsoon winds would be expected to alter the intensity of the 'freshwater plug'. Greater amounts of freshwater expelled from the Java Sea into the southern Indonesian seas would induce a colder ITF (lesser amounts would lead to a warmer ITF). As the regional precipitation changes with the phase of ENSO, Asian monsoon and at longer temporal scales, the ITF temperature may adjust accordingly. A strong 'freshwater plug' would reduce transfer of the western Pacific's warm pool into the Indian Ocean, with potential feedback to ENSO. The Indian Ocean surface layer would receive less Pacific heat, affecting its surface temperature and altering the pattern of heat and water vapour exchange with the atmosphere, presumably influencing the

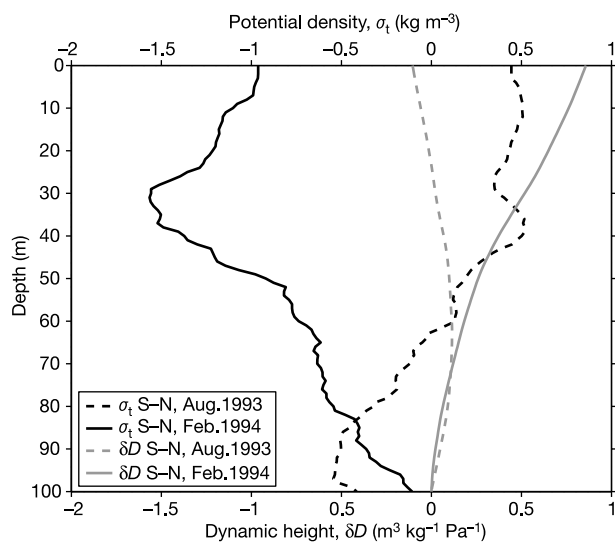


Figure 3 Differences in density and dynamic height within the upper 100 m between the northern and southern Makassar Strait (northern subtracted from southern) for two monsoon phases. In February 1994, (solid black line) representing the northwest monsoon, the density difference from the sea surface to a depth of 100 m is negative, meaning that the southern density profile is less dense or more buoyant than the northern profile. During August 1993 (solid grey line), representing the southeast monsoon, the density difference along the Makassar Strait is relatively small but positive in the upper 60 m. The dynamic height anomalies of the isobaric surfaces from 0 to 100 decibars (1 db is approximately equivalent to 1 m) relative to 100 db are shown as dashed lines. During the northwest monsoon (dashed black line), the surface layer isobaric surfaces in the southern Makassar Strait are higher than they are in the north, signifying a northward pressure gradient within the surface layer. During the southeast monsoon (dashed grey line), the meridional pressure gradient within the surface layer is essentially zero.

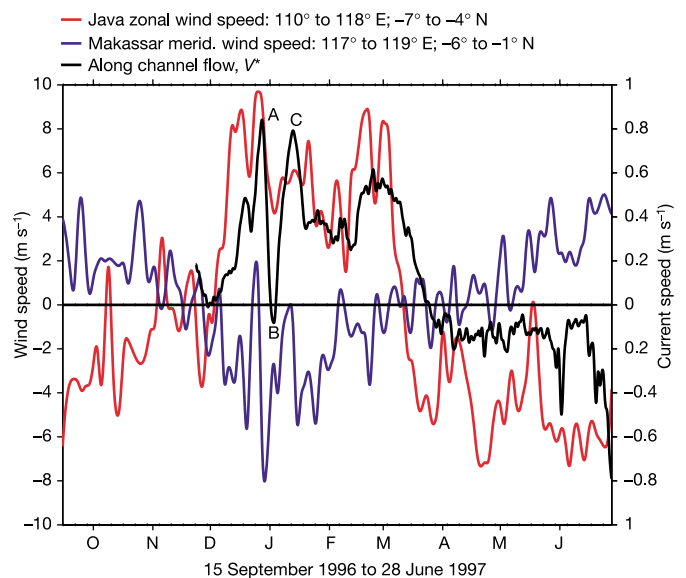


Figure 4 Makassar Strait surface layer current and Java and Makassar wind time series. The along-channel 'relative' surface layer current in the Makassar Strait (V^* , solid black line) is shown for the period of the ADCP time series. The values are daily averages; the 'relative' surface current curve is produced by subtracting the meridional current as measured within the thermocline by a current meter near 200 m from the average meridional current of the upper 50 m of the water column as measured by the ADCP. This removes the larger-scale ITF component from the more locally forced surface layer flow. The zonal wind in Java Sea (red line) and the meridional wind in Makassar Strait (blue line) reveal the close relationship of the wind of these two areas to the surface current. The wind data are derived from two-day-averaged wind as measured by the NSCAT satellite. The letters A, B and C denote the three events discussed in the text.

Asian monsoon and the Indian Ocean dipole^{25,26}

Models have been used to investigate the regional effect of varying the magnitude of the ITF on the Indian Ocean upper layer heat content and sea surface temperature^{5–9}. Changes in the sea surface temperature associated with changes in the ITF transport shift the position of the deep atmosphere convection region of the western tropical Pacific^{8,27} and, by changing the sea surface temperature in the Indian Ocean, alter the net evaporation within the Indian Ocean with consequences for the monsoon^{7,8,28}. Although models investigate the contrast between ‘off’ and ‘on’ ITF^{5–9}, they do not explicitly consider the effects or causes of varying the ITF transport and temperature profiles, yet changes in their relationship, even without changes in the net transport, may be expected to alter the sea surface temperature and heat budget of the Pacific and Indian oceans and associated climate phenomena. □

Received 15 January; accepted 8 September 2003; doi:10.1038/nature02038.

- Gordon, A. L. in *Ocean Circulation and Climate: Observing and Modeling the Global Ocean* (ed. Gould, J.) 303–314 (Academic, San Diego, 2001).
- Gordon, A. L. & Fine, R. A. Pathways of water between the Pacific and Indian Oceans in the Indonesian seas. *Nature* **379**, 146–149 (1996).
- Ffield, A., Vranes, K., Gordon, A. L., Susanto, R. D. & Garzoli, S. L. Temperature variability within Makassar Strait. *Geophys. Res. Lett.* **27**, 237–240 (2000).
- Vranes, K., Gordon, A. L. & Ffield, A. The heat transport of the Indonesian Throughflow and implications for the Indian Ocean heat budget. *Deep-Sea Res. II* **49**, 1391–1410 (2002).
- Lee, T., Fukumori, I., Menemenlis, D., Xing, Z. & Fu, L. Effects of the Indonesian Throughflow on the Pacific and Indian Oceans. *J. Phys. Oceanogr.* **32**, 1404–1429 (2002).
- Hirst, A. C. & Godfrey, J. S. The role of Indonesian Throughflow in a global ocean GCM. *J. Phys. Oceanogr.* **23**, 1057–1086 (1993).
- Godfrey, S. The effect of the Indonesian Throughflow on ocean circulation and heat exchange with the atmosphere: A review. *J. Geophys. Res.* **101**, 12217–12237 (1996).
- Wajsozowicz, R. Air-sea interaction over the Indian Ocean due to variations in the Indonesian Throughflow. *Clim. Dyn.* **18**, 437–453 (2002).
- Wajsozowicz, R. & Schneider, E. K. The Indonesian throughflow's effect on global climate determined from the COLA coupled climate system. *J. Clim.* **14**, 3029–3042 (2001).
- Ilahude, A. G. & Gordon, A. L. Thermocline stratification within the Indonesian Seas. *J. Geophys. Res.* **101**, 12401–12409 (1996).
- Van Aken, H. M., Punjawan, J. & Saimima, S. Physical aspects of the flushing of the east Indonesian basins. *Neth. J. Sea Res.* **22**, 315–339 (1988).
- Piola, A. R. & Gordon, A. L. On oceanic heat and fresh-water fluxes at 30°S. *J. Phys. Oceanogr.* **16**, 2184–2190 (1986).
- Toole, J. M. & Warren, B. A. A hydrographic section across the subtropical South Indian Ocean. *Deep-Sea Res. I* **40**, 1973–2019 (1993).
- Robbins, P. E. & Toole, J. M. The dissolved silica budget as a constraint on the meridional overturning circulation of the Indian Ocean. *Deep-Sea Res. I* **44**, 879–906 (1997).
- Macdonald, A. M. Property fluxes at 30S and their implications for the Pacific-Indian throughflow and the global heat budget. *J. Geophys. Res.* **98**, 6851–6868 (1993).
- Ganachaud, A. & Wunsch, C. Improved estimates of global ocean circulation, heat transport and mixing from hydrographic data. *Nature* **408**, 453–457 (2000).
- Ganachaud, A. & Wunsch, C. Large-scale ocean heat and freshwater transports during the World Ocean Circulation Experiment. *J. Clim.* **16**, 696–705 (2003).
- Gordon, A. L., Susanto, R. D. & Ffield, A. Throughflow within Makassar Strait. *Geophys. Res. Lett.* **26**, 3325–3328 (1999).
- Molcard, R., Fieux, M. & Ilahude, A. G. The Indo-Pacific throughflow in the Timor Passage. *J. Geophys. Res.* **101**, 12411–12420 (1996).
- Molcard, R., Fieux, M. & Syamsudin, F. The throughflow within Ombai Strait. *Deep-Sea Res. I* **48**, 1237–1253 (2001).
- Murray, S. P. & Arief, D. Throughflow into the Indian Ocean through the Lombok Strait, January 1985–January 1986. *Nature* **333**, 444–447 (1988).
- Oberhuber, J. M. *An Atlas Based on “COADS” Data Set* (Tech. Rep. 15, Max-Planck-Institut für Meteorologie, Hamburg, 1988).
- Wyrtki, K. *Physical Oceanography of the Southeast Asian Waters* (NAGA Rep. 2, Scripps Institution of Oceanography, La Jolla, 1961).
- Mariano, A. J., Ryan, E. H., Perkins, B. D. & Smithers, S. *The Mariano Global Surface Velocity Analysis* 55 (United States Coast Guard, Washington DC, 1995).
- Saji, N. H., Goswami, B. N., Vinayachandran, P. N. & Yamagata, T. A dipole mode on the tropical Indian Ocean. *Nature* **401**, 360–363 (1999).
- Susanto, R. D., Gordon, A. L. & Zheng, Q. N. Upwelling along the coasts of Java and Sumatra and its relation to ENSO. *Geophys. Res. Lett.* **28**, 1599–1602 (2001).
- Schneider, N. The Indonesian throughflow and the global climate system. *J. Clim.* **11**, 676–689 (1998).
- Wajsozowicz, R. C. & Schopf, P. S. Oceanic influences on the seasonal cycle in evaporation over the Indian Ocean. *J. Clim.* **14**, 1199–1226 (2001).
- Conkright, M. E. et al. *World Ocean Atlas CD-ROM Data Set Documentation* (Internal Report 15, National Oceanographic Data Center, Silver Spring, Maryland, 1998).

Acknowledgements This research is supported by the National Science Foundation, the National Aeronautics and Space Administration and the Office of Naval Research.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for material should be addressed to A.L.G. (agordon@ldeo.columbia.edu).

Independent rate and temporal coding in hippocampal pyramidal cells

John Huxter^{1*}, Neil Burgess^{1,2} & John O’Keefe^{1,2}

¹Department of Anatomy and Developmental Biology, ²Institute of Cognitive Neuroscience, University College London, Gower Street, London WC1E 6BT, UK

* Present address: Department of Anatomy, University of Bristol, Bristol BS8 1TD, UK

In the brain, hippocampal pyramidal cells use temporal¹ as well as rate² coding to signal spatial aspects of the animal’s environment or behaviour. The temporal code takes the form of a phase relationship to the concurrent cycle of the hippocampal electroencephalogram theta rhythm¹. These two codes could each represent a different variable^{3,4}. However, this requires the rate and phase to vary independently, in contrast to recent suggestions^{5,6} that they are tightly coupled, both reflecting the amplitude of the cell’s input. Here we show that the time of firing and firing rate are dissociable, and can represent two independent variables: respectively the animal’s location within the place field, and its speed of movement through the field. Independent encoding of location together with actions and stimuli occurring there may help to explain the dual roles of the hippocampus in spatial and episodic memory^{7,8}, or may indicate a more general role of the hippocampus in relational/declarative memory^{9,10}.

A cell must fire to manifest either a temporal or a rate code. Place cells are hippocampal pyramidal cells that increase their firing rate in a particular portion of the environment^{2,11} (the ‘place field’, Fig. 1b). As such, they provide a coarse rate code for the animal’s location within which a temporal code provides additional information^{1,12,13}. In addition, we propose that the rate of firing within the field can vary to encode other information without disrupting this temporal code.

We recorded the firing of place cells and the electroencephalogram (EEG) from the hippocampi of rats as they ran back and forth on a linear track for food reward at each end (Fig. 1a, Methods). During this behaviour, the EEG shows the prominent theta oscillation and each place cell fires in a specific region of the track (Fig. 1b). The cell’s bursting rate through the field (Fig. 1c) is slightly higher than the concurrent EEG theta frequency, so that the average phase of firing moves earlier on each theta wave as the animal progresses through the field (Fig. 1d). The phase of firing correlates with spatial variables such as the animal’s position on the track or within the field (Fig. 1d), and also with non-spatial variables such as the time since entry into the field (Fig. 1e) or instantaneous firing rate (IFR; Fig. 1f), but in general the correlation with position is stronger than with either (Fig. 1g; see also refs 1, 5, 6, 12). Some process must align the phase of each spike relative to the concurrent theta wave so that the phase codes for location despite the different speeds of each run through the field.

Must the phase and rate codes always co-vary with each other, or

is it possible for them to dissociate? Harris *et al.*⁵ reported an overall relationship between phase and rate in spikes recorded on a linear track, and our data show a similar relationship (Fig. 1g; Supplementary Fig. 1a, b). However, it is much weaker than the relationship between phase and position (Fig. 1d; phase correlates better with position than rate in 66/77 cells, $P = 5.3 \times 10^{-11}$), and may result directly from the relationship that both phase and rate hold with position (Supplementary Fig. 1c). Mehta and colleagues⁶ reported that, over the first few runs of a session, the firing fields became more negatively skewed versus position and the phase precession strengthened—suggesting that both rate and phase reflect the net input to the cell and that this input becomes ramp-shaped, increasing with distance through the field. Our fields show a continuous range of negative, zero and positive skew, but there was no correlation between skew and the rate or amount of phase precession (Supplementary Fig. 2). Thus a causal relationship between negative field skew and phase precession is unlikely, despite both effects strengthening over the first few runs of a session.

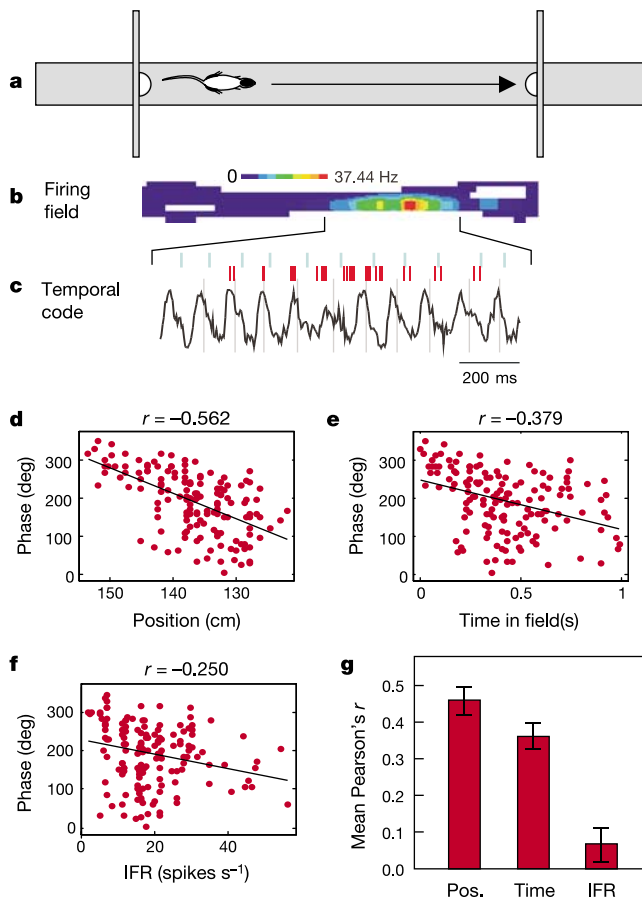


Figure 1 Place cell phase of firing correlates best with position. **a**, Behavioural task: rat shuttles back and forth along linear track between food rewards contained in cups attached to movable walls. **b**, False-colour firing field of a place cell created from multiple runs in the eastward direction. **c**, EEG theta rhythm and place cell firing (in red) for the same cell on a single eastward run. Ticks above the spikes indicate + to - zero crossings ($0^\circ/360^\circ$ phase) for each theta wave, lines through theta waves indicate 270° . Bursts of spikes occur at higher than theta frequency causing each successive burst to move to an earlier phase of the theta cycle, despite initially rising, then falling firing rate. Theta cycle phase of spikes for multiple runs from a place cell is plotted against position (**d**), time (**e**) and instantaneous firing rate (IFR; **f**) in the place field. **g**, Phase (adjusted for circularity, see Methods) is better correlated with location than with time or firing rate across the population of cells. Here and in subsequent figures, vertical bars represent $\pm$ s.e.m.

To see whether phase and rate dissociate within a place field, each field was divided into three equal segments: the beginning, middle and end. The mean IFR and mean firing phase per theta cycle, and their respective temporal derivatives (TD_{IFR} and TD_{phase}) for runs through each segment of the field were averaged across the population (Fig. 2). The mean phase per cycle continues to precess throughout the entire run, despite the firing rate rising in the early part of the field and then falling, and despite the increasing variance of firing phase through the field^{6,14}. TD_{IFR} is positive in the early part of the field and negative in the late part, while TD_{phase} is negative throughout, demonstrating that phase precession occurs during both accelerating and decelerating spike trains, and that the firing rate rises and falls within each run. Thus, again, a causal relationship between phase precession and increasing rate is unlikely, consistent with the much lower correlation of phase with firing rate than with position (Fig. 1).

To see whether rate and phase dissociate on a run-by-run basis, the runs with the highest and lowest firing rates for each cell were identified so that the phase precession in both data sets could be compared. Across the population, we found no difference in mean phase precession between the high- and low-firing-rate runs. Figure 3 shows that phase precession takes place equally on trials with low as well as high firing rates, and even under conditions of very low firing rates with two or fewer spikes per run (Fig. 3b). Thus the dissociation of firing rate and firing phase is not due to effects specific to the second part of the field such as habituation, spike frequency accommodation, or to high- and low-rate runs being combined in the overall mean rate. In addition, the above data rule out any necessary coupling between phase precession and TD_{IFR} (compare ref. 5).

If the phase and firing rate can be independent, what variables

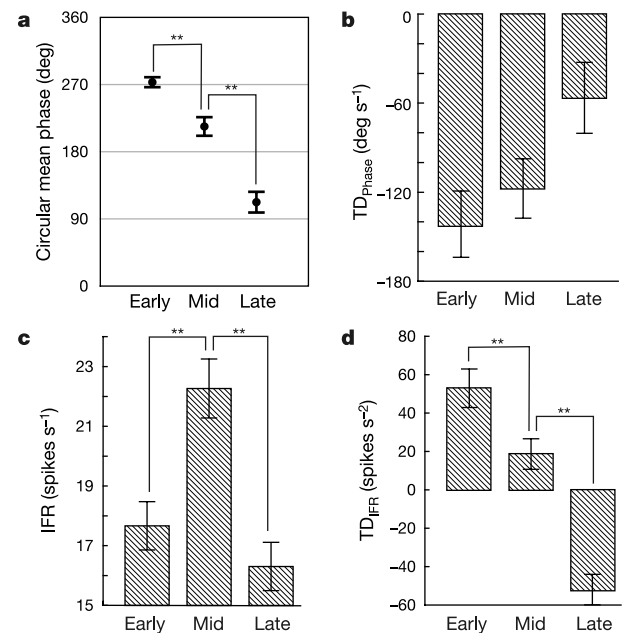


Figure 2 Phase precession is independent of IFR. **a**, Phase depends on location, being highest in the early third of each field, lower in the middle third, and lowest in the late third. **b**, Temporal derivative of phase (TD_{phase}) is negative in each portion of the field (68/76 fields in the early portion, $P < 1 \times 10^{-12}$; 57/76 in the middle, $P < 1 \times 10^{-5}$; 46/76 late, $P < 0.05$, binomial test). **c**, IFR starts low, increases in the middle third and then decreases in the late part of the field. **d**, Temporal derivative of instantaneous firing rate (TD_{IFR}) starts high, falls towards zero in the middle third, and then goes negative in the last third. Here and in subsequent figures, an asterisk denotes $P < 0.05$, and a double asterisk denotes $P < 0.01$.

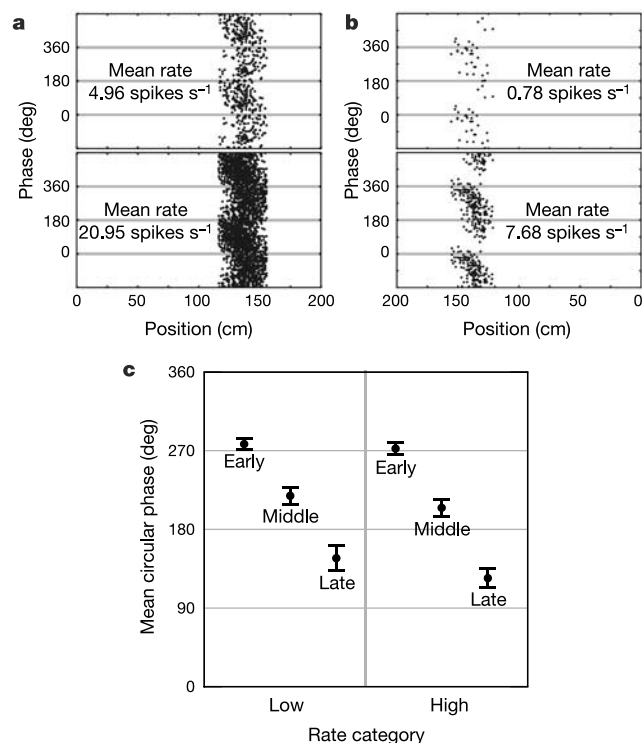


Figure 3 Phase is correlated with track location on low- as well as high-firing-rate runs. **a**, A single cell for which average firing rate on high-rate runs (lower panel) is four times that on lower rate ones (upper panel) with no discernible effect on the phase precession. **b**, The same analysis for the cell shown in Fig. 1d–f. Phase precession occurs despite very low firing rates. Above: 35 spikes from 66 low-rate runs, mean spikes per run 0.53, s.e.m. 0.09, range 0–2 (0 spikes on 41 runs, 1 on 15, and 2 on 10); below: 194 spikes from 36 high-rate runs, mean spikes per run 5.40, s.e.m. 0.53, range 2–14. **c**, Population data ($n = 34$): phase precession across early, middle and late parts of the field on both high- and low-rate runs.

does each encode? Previous work has suggested that the firing rate of place cells correlates with running speed through the field^{15–18}. We confirm that relationship for the present data set (Fig. 4), where place fields reconstructed from faster runs were associated with higher firing rates (30 of 34 fields, $t_{1,33} = -5.88$, $P < 0.001$; Fig. 4a). In contrast, the temporal code did not vary between fast and slow runs, as measured by mean rate of phase precession in space ($t_{1,33} = 0.301$, $P = 0.77$; Fig. 4b); total extent of phase precession ($t_{1,33} = -4.32$, $P = 0.67$; Fig. 4c); phase at firing onset ($F_{1,32} = 0.066$, $P = 0.799$) or offset ($F_{1,32} = 1.281$, $P = 0.266$; data not shown). The correlation between mean firing rate per run and running speed on individual runs, averaged across the population, is 0.223 (Fig. 4d)—higher than the correlation between speed and any other variable that we have measured.

To identify the spatial correlate of firing phase more precisely, we varied the length of the track from 1.5 m to 1.0 m or 0.75 m by moving the end walls closer together. The rats ran slower on the shortened tracks ($F_{1,29} = 12.536$, $P = 0.001$), the fields were smaller ($F_{1,29} = 16.834$, $P < 0.001$) and firing rates were reduced ($F_{1,29} = 11.068$, $P = 0.002$) (Fig. 5a). When the fields shrank, there was a strong correlation between the shortening of a field and the increase in the rate of change of phase in space (Fig. 5b). By contrast, the change in the rate of phase precession was not correlated with changes in its firing rate (Fig. 5c). We tentatively conclude that the phase is coding for the proportion of the field traversed, which is also consistent with the finding that blocking long-term potentiation (LTP) produces shorter fields on average (by

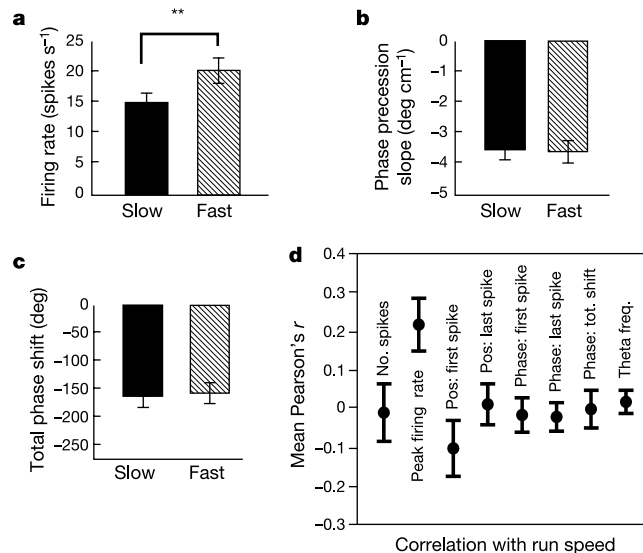


Figure 4 Firing rates differ on fast and slow runs through the field. **a**, Significant difference in average peak firing rates but no difference in **b**, rate of phase precession, or **c**, total phase shift. **d**, Firing rate is the only variable measured that consistently correlates with running speed (1.5 m track, mean $r = 0.223$, 23/29 fields with significant positive correlations $P < 0.05$).

preventing their asymmetric expansion¹⁹) but does not reduce the amount of phase precession¹⁸. The correlation between running speed and firing rate per run observed on the full-length track was also found on the shortened track (mean $r = 0.18$), showing the robustness of this effect within different environmental configurations.

What mechanisms might underlie phase coding of location independent of rate? One model^{1,20} views CA3 place-cell firing as an interference pattern between an inhibitory theta input and an intrinsically oscillatory membrane potential (see, for example, ref. 21), whose frequency increases above theta frequency as the input to the cell increases (see, for example, ref. 22). Under this model, modulation of the amplitudes of the response, possibly by input from the dentate gyrus, need not affect its phase, while the rate of precession naturally adjusts to cover the extent of the field. If the frequency of the variable oscillator reflects distance from landmarks as measured by external and internal signals including motor efference, this model could also account for the phase precession observed in running wheels as well as linear tracks⁵. Alternative possibilities include phase of firing reflecting the directions of the stimuli defining the place field^{3,23}, or early-field firing being driven by input from other place cells while late-field firing is driven by sensory input^{19,24–26}. In this latter model, delays in synaptic transmission mean that early-field firing occurs later in a theta cycle than that driven by sensory input (note, however, that the drive from other place cells should not depend on LTP¹⁸).

The present experiments show that, on linear tracks, the phase of firing relative to the hippocampal EEG theta wave and the IFR code for two independent aspects of the animal's spatial behaviour: proportion of distance through the place field, and in-field velocity. But the correlation of rate with velocity varies between cells and is low overall, suggesting that it might also code for other variables. Previous experiments^{2,16,27,28} suggest that firing rate in the place field is influenced by such variables as direction of turning along the rat's trajectory and the presence (or absence) of particular objects, or of smells or sounds associated with reinforcers. Thus we propose that the strong spatial relation with phase leaves modulation of the firing rate free to represent additional information regarding the beha-

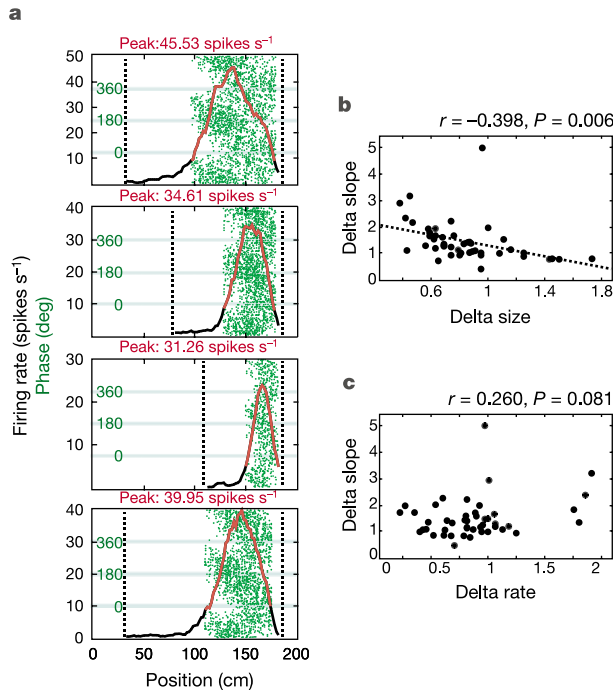


Figure 5 Phase precession on the shortened tracks. Single cell: **a**, phase angle versus position (green) gets steeper as the field size shrinks and the field firing rate (spikes dividing by dwell time, in red) drops owing to shortening of the track. Bottom panel shows the second baseline trial on full-length track carried out after short track trials. Population: slope of phase precession (that is rate of change in space) increases with shorter track length (**b**) in the absence of systematic changes in other correlations such as firing rate (**c**). 'Delta' values refer to the change from the preceding baseline trial to the trial on the shortened track (see Methods).

viours performed and the salient objects or features encountered in specific places. This dual code may provide the neural basis for the involvement of the hippocampus in both spatial and more general episodic/declarative memory^{7–10} by simultaneously encoding and binding together the location of the event and its behavioural and sensory content. Theta EEG activity has also been demonstrated in many different neocortical regions in human patients²⁹, suggesting that the ability to code for more than one variable by using rate and temporal codes may be a general property of cortical pyramidal cells. □

Methods

Experimental procedure

Nine male Lister hooded rats were implanted with pairs of movable tetrodes in the dorsal hippocampus^{1,30}, and placed on a food deprivation schedule (initially reduced to 90% of body weight with a subsequent gain of 5 g week⁻¹). After one week of post-operative recovery, they were trained to shuttle between the ends of a linear track for food reward (track dimensions: 150 × 10 cm, bounded by 25 cm side walls and 61 cm end walls that could be moved to shorten the length of the track). Above the walls, the rats could see the experimental room. A recording session consisted of a series of 8-min trials, the first always a set of runs on the full-length track followed by a series of trials including trials on shortened tracks, interspersed with baseline full-length track trials. Further series were then run in which the size of the end walls was modified or the belt which comprised the floor of the track was moved. Only the results from the first series are reported here.

An overhead camera tracked an array of infrared light-emitting diodes (LEDs) fixed to the rat's head (sampled at 50 Hz with a theoretical resolution of 3 mm). Simultaneously, EEG field potentials (sampled at 250 Hz, band-pass filtered at 0.34–125 Hz) and extracellular action potentials (spikes, sampled at 48 kHz, band-pass filtered at 500–6,700 Hz) were recorded from the pyramidal cell layer. Half-sinusoids in the theta frequency range (6–16 Hz) were aligned to negative-going deflections in the EEG and the goodness-of-fit used to detect the presence of theta. The firing phase (0–359°) of a spike reflects its temporal distance between successive zero crossings during theta. Where absolute phase values from different fields are combined, each field is first adjusted by a constant so that least frequent phase corresponds to 0°/360°.

Place field definition

For a given cell, firing rates were calculated in 2 cm bins along the length of the track (that is, the ratio of the total number of spikes to occupancy duration in each bin) and then boxcar-averaged using the five bins centred on it. Place fields were defined as the bin with the highest firing rate and all contiguous bins in which the firing rate exceeded 20% of this peak rate.

General data analysis

Only well-isolated units with good spatial selectivity and a peak firing rate of >1 Hz were analysed, yielding 73 cells with 94 place fields: 75 from CA1, 19 from CA3, the dentate gyrus or hilar region. All further analyses concerned data from runs through the place field in its preferred firing direction at above 10 cm s⁻¹. Positions are plotted such that the ordinate increases for runs from east to west and decreases for runs from west to east. The initial trial on the full-length track from each place field was used to compare field characteristics such as peak firing rate, field skew, and rate of phase precession versus position. Correlations and regressions on phase data were adjusted to take account of its circular nature, minimizing the squared angular distance from the line rather than the squared absolute distance (see also ref. 1). Correlations were done on the 77 fields contributing at least 100 spikes to the initial full-length trial.

IFR was calculated as the number of spikes in a time window up to one theta cycle on either side of the reference spike, divided by the size of the window (after ref. 5). The temporal derivative of the IFR (TD_{IFR}) is the difference between the IFR for any two consecutive spikes, divided by the time interval between them. Only IFRs based on time windows greater than 100 ms were included. Mean phase per cycle is the circular mean phase of spikes within a theta cycle, and the temporal derivative of this (TD_{phase}) is the circular difference between the mean phase in successive theta cycles divided by the duration of the first cycle.

For analyses comparing fast versus slow runs and high-rate versus low-rate runs (Figs 3c and 4a–c), separate fields were constructed from subsets of runs (combining multiple baseline trials with $r > 0.9$ field correlations). 'Slow runs' comprised those containing the bottom quartile of spikes ordered by the speed of the run through the firing field (that is, the length of the path through the field divided by its duration), 'fast runs' comprised the top quartile. Four fields for which this produced fewer than 200 spikes in each subset were simply split at the 50th percentile. Runs ordered by firing rate were divided into 'high' and 'low' rate runs at the 50th percentile of the firing rate. In both cases, only fields for which there were at least 100 runs through the composite field were included ($n = 34$).

Shortened runway

Field characteristics on two shortened-runway trials (created by moving one of the end walls 50 or 75 cm towards the middle of the track) were compared with the two flanking baseline trials and analysed using a repeated measures analysis of variance. The wall that was moved and the amount by which it was moved varied pseudo-randomly across trials. Changes in firing rate, running speed, and field size were measured as the ratio of the score on each manipulation trial compared to the preceding baseline. Changes in the rate of precession were measured as the difference between scores on manipulation trials and the preceding baseline, and were only performed on trials with 100 or more spikes.

Received 15 May; accepted 15 September 2003; doi:10.1038/nature02058.

- O'Keefe, J. & Recce, M. L. Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus* **3**, 317–330 (1993).
- O'Keefe, J. Place units in the hippocampus of the freely moving rat. *Exp. Neurol.* **51**, 78–109 (1976).
- O'Keefe, J. in *Brain and Space* (ed. Paillard, J.) 273–295 (Oxford Univ. Press, Oxford, 1991).
- Buzsáki, G. Theta oscillations in the hippocampus. *Neuron* **33**, 325–340 (2002).
- Harris, K. D. et al. Spike train dynamics predicts theta-related phase precession in hippocampal pyramidal cells. *Nature* **417**, 738–741 (2002).
- Mehta, M. R., Lee, A. K. & Wilson, M. A. Role of experience and oscillations in transforming a rate code into a temporal code. *Nature* **417**, 741–746 (2002).
- O'Keefe, J. & Nadel, L. *The Hippocampus as a Cognitive Map* (Oxford Univ. Press, Oxford, 1978); (www.cognitivemap.net) (2003).
- Burgess, N., Maguire, E. A. & O'Keefe, J. The human hippocampus and spatial and episodic memory. *Neuron* **35**, 625–641 (2002).
- Squire, L. R. Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychol. Rev.* **99**, 195–231 (1992).
- Eichenbaum, H. & Cohen, N. *From Conditioning to Conscious Recollection* (Oxford Univ. Press, Oxford, 2001).
- Muller, R. A quarter of a century of place cells. *Neuron* **17**, 813–822 (1996).
- Skaggs, W. E., McNaughton, B. L., Wilson, M. A. & Barnes, C. A. Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences. *Hippocampus* **6**, 149–172 (1996).
- Jensen, O. & Lisman, J. E. Position reconstruction from an ensemble of hippocampal place cells: Contribution of theta phase coding. *J. Neurophysiol.* **83**, 2602–2609 (2000).
- Yamaguchi, Y., Aota, Y., McNaughton, B. L. & Lipa, P. Bimodality of theta phase precession in hippocampal place cells in freely running rats. *J. Neurophysiol.* **87**, 2629–2642 (2002).
- McNaughton, B. L., Barnes, C. A. & O'Keefe, J. The contributions of position, direction, and velocity to single unit activity in the hippocampus of freely-moving rats. *Exp. Brain Res.* **52**, 41–49 (1983).
- Wiener, S. I., Paul, C. A. & Eichenbaum, H. Spatial and behavioral correlates of hippocampal neuronal activity. *J. Neurosci.* **9**, 2737–2763 (1989).
- Hirase, H., Czurko, H. H., Csicsvari, J. & Buzsáki, G. Firing rate and theta-phase coding by hippocampal pyramidal neurons during 'space clamping'. *Eur. J. Neurosci.* **11**, 4373–4380 (1999).
- Ekstrom, A. D., Meltzer, J., McNaughton, B. L. & Barnes, C. A. NMDA receptor antagonism blocks experience-dependent expansion of hippocampal "place fields". *Neuron* **31**, 631–638 (2001).
- Mehta, M. R., Barnes, C. A. & McNaughton, B. L. Experience-dependent, asymmetric expansion of hippocampal place fields. *Proc. Natl Acad. Sci. USA* **94**, 8918–8921 (1997).

20. Lengyel, M., Szatmary, Z. & Erdi, P. Dynamically detuned oscillators account for the coupled rate and temporal code of place cell firing. *Hippocampus* **13**, 700–714 (2003).
21. Pike, F. G. *et al.* Distinct frequency preferences of different types of rat hippocampal neurones in response to oscillatory input currents. *J. Physiol. Lond.* **529**, 205–213 (2000).
22. Kamondi, A., Acsády, L., Wang, X. J. & Buzsáki, G. Theta oscillations in somata and dendrites of hippocampal pyramidal cells *in vivo*: Activity-dependent phase-precession of action potentials. *Hippocampus* **8**, 244–261 (1998).
23. Burgess, N., Recce, M. & O'Keefe, J. A model of hippocampal function. *Neural Netw.* **7**, 1065–1081 (1994).
24. Tsodyks, M. V., Skaggs, W. E., Sejnowski, T. J. & McNaughton, B. L. Population dynamics and theta rhythm phase precession of hippocampal place cell firing: A spiking neuron model. *Hippocampus* **6**, 271–280 (1996).
25. Jensen, O. & Lisman, J. E. Hippocampal CA3 region predicts memory sequences: Accounting for the phase precession of place cells. *Learn. Mem.* **3**, 279–287 (1996).
26. Wallenstein, G. V. & Hasselmo, M. E. GABAergic modulation of hippocampal population activity: Sequence learning, place field development, and the phase precession effect. *J. Neurophysiol.* **78**, 393–408 (1997).
27. Wood, E. R., Dudchenko, P. A. & Eichenbaum, H. The global record of memory in hippocampal neuronal activity. *Nature* **397**, 613–616 (1999).
28. Moita, M. A., Rosis, S., Zhou, Y., LeDoux, J. E. & Blair, H. T. Hippocampal place cells acquire location-specific responses to the conditioned stimulus during auditory fear conditioning. *Neuron* **37**, 485–497 (2003).
29. Kahana, M. J., Sekuler, R., Caplan, J. B., Kirschen, M. & Madsen, J. R. Human theta oscillations exhibit task dependence during virtual maze navigation. *Nature* **399**, 781–784 (1999).
30. O'Keefe, J. & Burgess, N. Geometric determinants of the place fields of hippocampal neurons. *Nature* **381**, 425–428 (1996).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank C. Lever, F. Cacucci, T. Wills, J. Ryan, D. Edwards and T. Hartley for technical assistance. This work was supported by the MRC and the Wellcome Trust. J.H. was a Rothermere Fellow.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.O'K. (j.okeefe@ucl.ac.uk).

A regulatory mutation in *IGF2* causes a major QTL effect on muscle growth in the pig

Anne-Sophie Van Laere^{1*}, Minh Nguyen^{3*}, Martin Braunschweig¹, Carine Nezer³, Catherine Collette³, Laurence Moreau³, Alan L. Archibald⁴, Chris S. Haley¹, Nadine Buys⁵, Michael Tally⁶, Göran Andersson¹, Michel Georges³ & Leif Andersson^{1,2}

¹Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, and ²Department of Medical Biochemistry and Microbiology, Uppsala University, BMC, Box 597, SE-751 24 Uppsala, Sweden

³Department of Genetics, Faculty of Veterinary Medicine, University of Liege (B43), 20, bd. de Colonster, 4000 Liege, Belgium

⁴Roslin Institute (Edinburgh), Roslin, Midlothian EH25 9PS, Scotland, UK

⁵Gentec, Kapelbaan 15, 9255 Buggenhout, Belgium

⁶Tally Consulting, SE-11458 Stockholm, Sweden

* These authors contributed equally to this work

Most traits and disorders have a multifactorial background indicating that they are controlled by environmental factors as well as an unknown number of quantitative trait loci (QTLs)^{1,2}. The identification of mutations underlying QTLs is a challenge because each locus explains only a fraction of the phenotypic variation^{3,4}. A paternally expressed QTL affecting muscle growth, fat deposition and size of the heart in pigs maps to the *IGF2* (insulin-like growth factor 2) region^{5,6}. Here we show that this QTL is caused by a nucleotide substitution in intron 3 of *IGF2*. The mutation occurs in an evolutionarily conserved CpG island that is hypomethylated in skeletal muscle. The mutation

abrogates *in vitro* interaction with a nuclear factor, probably a repressor, and pigs inheriting the mutation from their sire have a threefold increase in *IGF2* messenger RNA expression in post-natal muscle. Our study establishes a causal relationship between a single-base-pair substitution in a non-coding region and a QTL effect. The result supports the long-held view that regulatory mutations are important for controlling phenotypic variation⁷.

The QTL affecting muscle growth, fat deposition and heart size was first identified in intercrosses between the European wild boar and Large White domestic pigs and between Piétrain and Large White pigs^{5,6}. The alleles from the Large White breed in the first cross and the Piétrain breed in the second cross increased muscle mass and reduced back-fat thickness. The QTL explained 15–30% of the phenotypic variation in muscle mass and 10–20% of the variation in back-fat thickness^{5,6}. We recently used a haplotype-sharing approach to refine the map position of the QTL⁸. We assumed that a new allele (*Q*) promoting muscle development occurred *g* generations ago on a chromosome carrying the wild-type allele (*q*). We also assumed that the favourable allele had gone through a selective sweep due to the strong selection for lean growth in commercial pig populations. A QTL genotype cannot be deduced directly from an individual's phenotype but the QTL genotype of sires can be determined by progeny testing and marker-assisted segregation analysis². Twenty-eight chromosomes with known QTL status were identified. All 19 *Q*-bearing chromosomes shared a haplotype in the 250-kilobase (kb) interval between the markers *370SNP6/15* and *SWC9* (*IGF2* 3' untranslated region), which was therefore predicted to contain the QTL. This region contains *INS* and *IGF2* as the only known paternally expressed genes. Given their known functions and especially the role of *IGF2* in myogenesis⁹, they stood out as prime positional candidates.

We re-sequenced one of the 19 *Q* chromosomes (P208) and six *q* chromosomes (each corresponding to a distinct marker haplotype) for a 28.6-kb segment containing *IGF2*, *INS* and the 3' end of *TH*. This chromosome collection was expanded by including *Q* and *q* chromosomes from the following: (1) a wild boar/Large White intercross segregating for the QTL⁵; (2) a Swedish Landrace boar showing no evidence for QTL segregation in a previous study¹⁰; and (3) *F*₁ sires from a Hampshire/Landrace cross and a Meishan/Large White intercross both showing no indication for QTL segregation (see Methods). The lack of evidence for QTL segregation shows that the boars are either homozygous *Q/Q* or *q/q*. A Japanese wild boar was included as a reference for the phylogenetic analysis and it was assumed to be homozygous wild type (*q/q*). We identified a total of 258 DNA sequence polymorphisms corresponding to one polymorphic nucleotide per 111 base pairs (bp) (Fig. 1). Two major and quite divergent clusters of haplotypes were revealed (Supplementary Fig. 1). The two established *Q* haplotypes from Piétrain and Large White animals (P208 and LW3) were identical to each other and to the chromosomes from the Landrace (LR) and Hampshire/Landrace (H205) animals, showing that the latter two must be of *Q* type as well. The absence of QTL segregation in the offspring of the *F*₁ Hampshire × Landrace boar carrying the H205 and H254 chromosomes implies that the latter recombinant chromosome is also of *Q* type. This places the causative mutation downstream from *IGF2* intron 1, in the region for which H254 is identical to the other *Q* chromosomes. The Large White chromosome (LW197) from the Meishan/Large White pedigree clearly clustered with *q* chromosomes, implying that the *F*₁ sire used for sequencing was homozygous *q/q* as no overall evidence for QTL segregation was observed in this intercross. This is consistent with a previous study showing that Meishan pigs carry an *IGF2* allele associated with low muscle mass¹¹. Surprisingly, the Meishan allele (M220) was nearly identical to the *Q* chromosomes but with one notable exception, it shared guanine with all *q* chromosomes at a position (*IGF2*-intron3-nucleotide 3072) where all *Q* chromosomes have adenine (Fig. 1).

Under a bi-allelic QTL model, the causative mutation would

correspond to a DNA polymorphism for which the two alleles segregate perfectly between *Q* and *q* chromosomes. The G to A transition at *IGF2*-intron3-3072 is the only polymorphism fulfilling this criterion, implying that it is the causative quantitative trait nucleotide (QTN)¹. We genotyped the founders and 12 *F*₁ boars for the putative QTN to verify the QTL status of animals from the Meishan/Large White intercross. All founders and *F*₁ boars were homozygous G/G (*q/q*) except one Large White founder and one *F*₁ boar, which were heterozygous A/G. A QTL analysis revealed that the heterozygous boar, but no other *F*₁ sire, showed clear evidence for segregation of a paternally expressed QTL, and the Meishan allele increased back-fat thickness as predicted (χ^2 with 1 degree of freedom = 7.75, $P = 0.005$). Including this, we have so far tested 13 large sire families where the sire is heterozygous A/G at the QTN, and all have shown evidence for QTL segregation. In contrast, we have tested more than 50 sires, representing several different breeds, genotyped as homozygous A/A or G/G without obtaining any evidence for the segregation of a paternally expressed QTL at the *IGF2* locus. The results provide conclusive genetic evidence that *IGF2*-intron3-G3072A is the causative mutation. The Meishan allele is apparently identical to the ancestral haplotype on which the mutation occurred.

IGF2-intron3-3072 is part of an evolutionarily conserved CpG island of unknown function¹² located between differentially methylated region 1 (DMR1) and a matrix attachment region previously

defined in mice^{13–15}. The 94-bp sequence around the mutation shows about 85% identity to human, and the wild-type nucleotide at *IGF2*-intron3-3072 is conserved among eight mammalian species (Fig. 1). The QTN occurs 3 bp downstream of a conserved 8-bp palindromic sequence. The methylation status of the 300-bp fragment centred on *IGF2*-intron3-3072 and containing 50 CpG dinucleotides was examined by bisulphite sequencing in four-month-old *Q*^{pat}/*q*^{mat} and *q*^{pat}/*Q*^{mat} pigs. The CpG island was methylated in liver (26% of CpGs methylated on average), whereas in skeletal muscle both paternal (pat) and maternal (mat) chromosomes were essentially unmethylated (including the *IGF2*-intron3-3071 C residue) irrespective of QTL genotype (3.4% of CpGs methylated on average, Fig. 2a; see also Supplementary Fig. 2).

To uncover a possible function for this element, we performed electrophoretic mobility shift assays (EMSAs) using wild-type (*q*) and mutant (*Q*) sequences. Nuclear extracts were incubated with radioactively labelled *q* or *Q* double-stranded oligonucleotides. One specific complex (C1 in Fig. 2b) was obtained with the unmethylated wild-type (*q*) but not the mutant (*Q*) probe using extracts from murine C2C12 myoblasts. This complex was not obtained with the *q*^{*} probe, which has a methylated CpG at the QTN (Fig. 2b). A complex with approximately the same migration—but slightly weaker—was also detected in extracts from human HEK293 embryonic kidney cells and HepG2 hepatocytes. The specificity of the complex was confirmed as competition was obtained with 10-

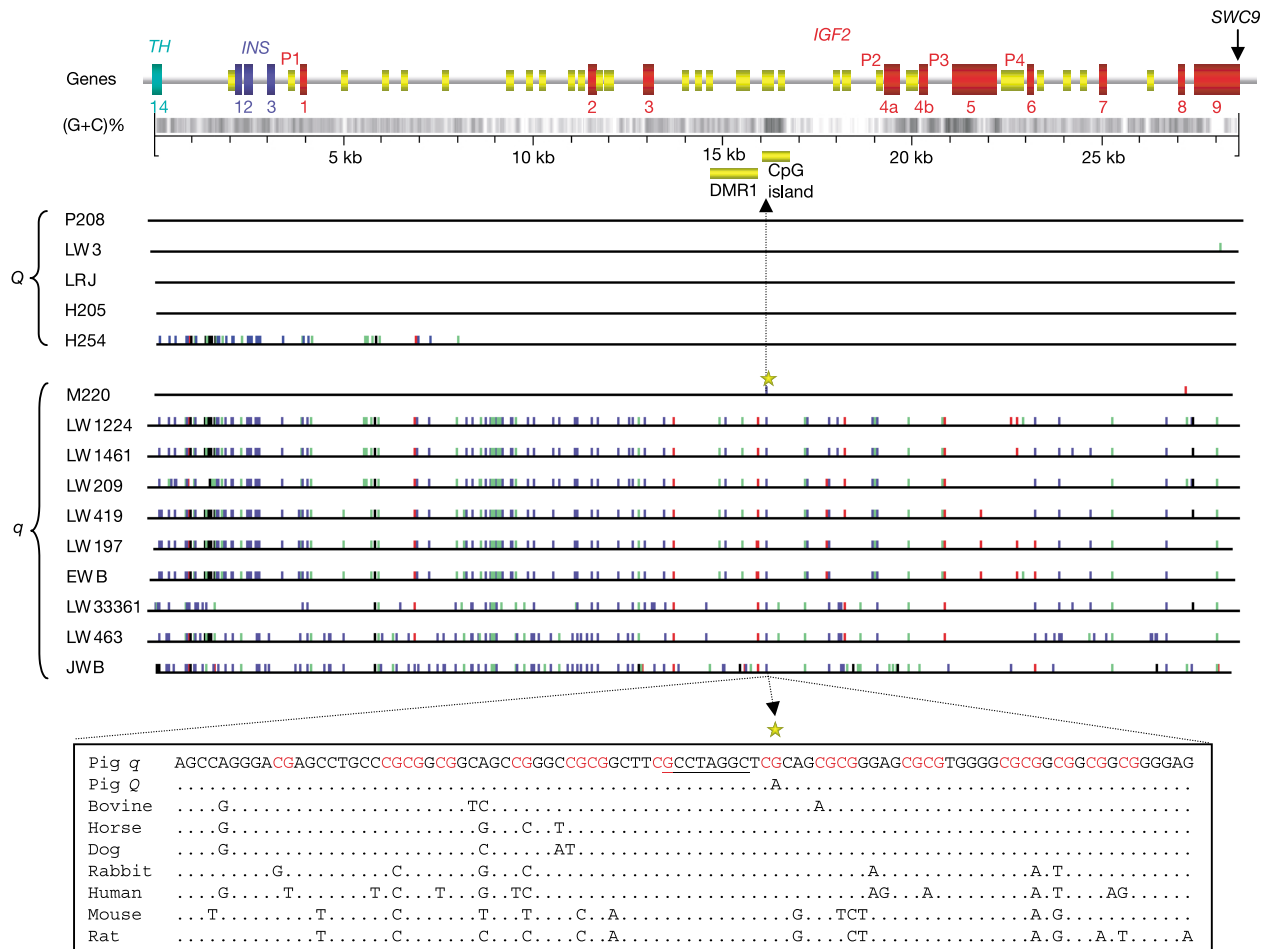


Figure 1 Polymorphisms in a 28.6-kb segment containing *TH* (exon 14), *INS* and *IGF2* among 15 pig chromosomes with deduced QTL status. The (G+C) content of a moving 100-bp window is shown on a grey scale (black 100%, white 0%). Yellow cylinders mark evolutionarily conserved regions¹². Viewgene²⁹ was used to highlight differences between the reference P208 sequence and other chromosomes. Blocks

indicate transitions (blue), transversions (green), insertions (red) and deletions (black). The QTN is indicated with an asterisk and the surrounding sequence is shown for eight mammals (CpGs are highlighted in red; a palindromic sequence is underlined). P, Piétrain; LW, Large White; LR, Landrace; H, Hampshire; M, Meishan; EWB, European wild boar; JWB, Japanese wild boar.

fold molar excess of unlabelled *q* probe, whereas a 50-fold excess of unlabelled *Q* probe or methylated *q** probe did not compete (Fig. 2b). Thus, the wild-type sequence binds a nuclear factor, and this interaction is abrogated by the mutation or methylation of the actual CpG site.

We analysed the effect of the *IGF2* *Q* mutation on transcription by using a transient transfection assay in mouse C2C12 myoblasts. We made *Q* and *q* constructs containing a 578-bp fragment from the actual region inserted in front of a luciferase reporter gene driven by the endogenous pig *IGF2* promoter 3 (*P3*) located approximately 5 kb downstream from the QTN (Fig. 1). This promoter was chosen because it generates the predominant *IGF2* transcript in muscle and the QTN affected the amount of *P3* transcripts *in vivo* (see below). The *q* fragment clearly acted as a repressor element and reduced

luciferase activity to about 25%, whereas the *Q* fragment was a significantly weaker repressor element and showed about 70% activity compared with *P3* alone (Fig. 2c). Our interpretation of this result, combined with those from the EMSA experiment, is that the *Q* mutation abrogates the interaction with a putative repressor protein. Thus, the *IGF2*-intron3-G3072A transition may be sufficient to explain the QTL effect as the two constructs only differ at this position. The constructs were also inserted in front of the heterologous herpes thymidine kinase (*TK*) minimal promoter. In this case the *q* construct caused a twofold increase of transcription whereas the *Q* construct caused a significantly higher, sevenfold increase (Supplementary Fig. 3). The results support our interpretation that the *q* allele represses transcriptional activity in C2C12 myoblasts when compared with the *Q* allele.

The *in vivo* effect of the mutation on *IGF2* expression was studied in a purpose-built *Q/q* × *Q/q* intercross. We tested the effect of the intron3-3072 mutation on *IGF2* imprinting, as a deletion encompassing DMR0, DMR1 and the associated CpG island derepresses the maternal *IGF2* allele in mesodermal tissues in mouse¹⁴. This was achieved by monitoring transcription from paternal and maternal alleles in tissues of *q/q*, *Q^{pat}/q^{mat}* and *q^{pat}/Q^{mat}* animals heterozygous for the *SWC9* microsatellite located in the *IGF2* 3' untranslated region. Before birth, *IGF2* was expressed exclusively from the paternal allele in skeletal muscle and kidney, irrespective of QTL genotype. At four months of age, weak expression from the maternal allele was observed in skeletal muscle, however at comparable rates for all three QTL genotypes (Supplementary Fig. 4). Only the paternal allele could be detected in four-month-old kidney. Consequently, the mutation does not seem to affect the

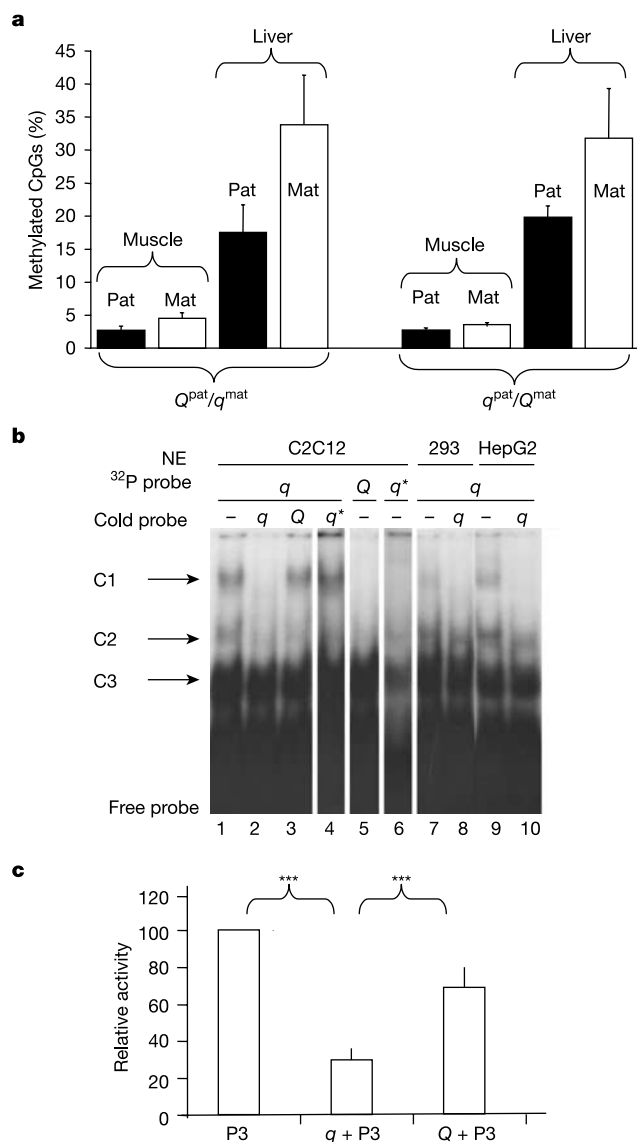


Figure 2 Methylation status, EMSA and transfection assays assessing the significance of the *IGF2* mutation. **a**, Percentage methylation around the QTN in liver and skeletal muscle of four-month-old pigs. Means ± s.e.m. are given; the numbers of analysed paternal (Pat) and maternal (Mat) chromosomes were in the range 10–26. **b**, EMSA using nuclear extracts (NE) from C2C12, HEK293 and HepG2 cells. Complex 1 (C1) was exclusively detected with the *q* probe. C2 was stronger in *q* but also probably present in the *Q* lane. C3 was unspecific. **c**, Luciferase assays of reporter constructs using pig *IGF2* *P3* promoter and intron 3 fragments (*Q* and *q*). Relative activities compared with the *P3*–LUC reporter are reported (means ± s.e.m.). Triple asterisk, *P* < 0.001.

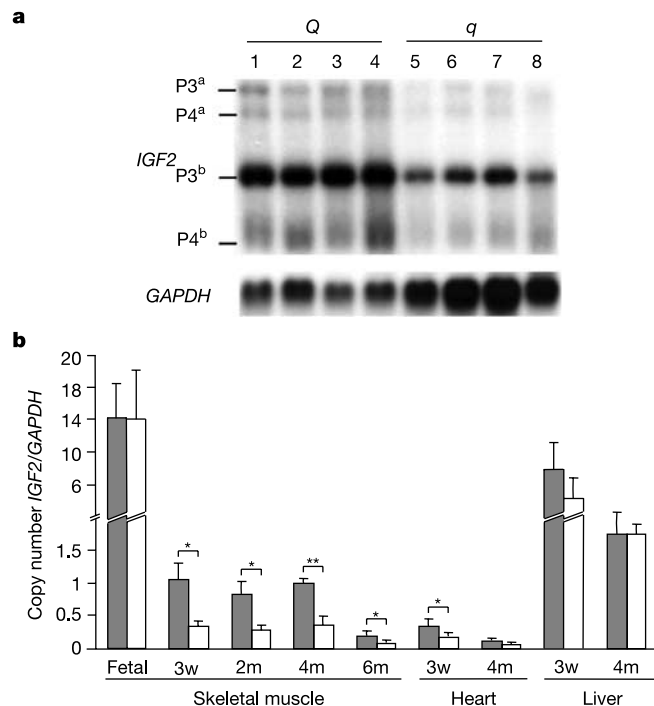


Figure 3 Analysis of *IGF2* mRNA expression. **a**, Northern blot of skeletal muscle (*gluteus*) poly(A)⁺ RNA from three-week-old piglets. Animals 1–4 and 5–8 carried a paternal *IGF2* *Q* or *q* allele, respectively. *P3* and *P4* indicate promoter usage, and *a/b* superscripts indicate the alternative polyadenylation signal used. All four transcripts showed a higher relative expression (standardized using *GAPDH*) in the *Q* group (*P* < 0.05). **b**, Real-time PCR analysis of *IGF2* expression in pigs carrying paternal *IGF2* *Q* (grey columns) or *q* (white columns) alleles. Expression levels were normalized using *GAPDH*. Means ± s.e.m. are given, *n* = 3–11. Asterisk, *P* < 0.05; double asterisk, *P* < 0.01; w, week; m, month.

imprinting status of *IGF2*. The partial derepression of the maternal allele in skeletal muscle may however explain why in a previous study muscle growth was found to be slightly superior in $q^{\text{pat}}/Q^{\text{mat}}$ versus q/q animals, and in Q/Q versus $Q^{\text{pat}}/q^{\text{mat}}$ animals⁵.

The *Q* allele was expected to be associated with increased *IGF2* expression because IGF-II stimulates myogenesis⁹. We monitored the relative mRNA expression of *IGF2* at different ages in the $Q/q \times Q/q$ intercross using both northern blot analysis and real-time polymerase chain reaction (PCR) (Fig. 3). The expression levels in fetal muscle and postnatal liver at three weeks of age were approximately tenfold higher compared with postnatal muscle. No significant difference was observed in fetal samples or in postnatal liver samples, but a significant threefold increase of postnatal *IGF2* mRNA expression in skeletal muscle was observed in Q/Q or $Q^{\text{pat}}/q^{\text{mat}}$ versus $q^{\text{pat}}/Q^{\text{mat}}$ or q/q progeny. There was also a significant, but less pronounced, increase of mRNA expression in heart associated with the Q^{pat} allele. The significant difference in *IGF2* expression revealed by real-time PCR was confirmed using two different internal controls: GAPDH (Fig. 3b) and HPRT (data not shown). We found an increase of all detected transcripts originating from the three promoters (P2–P4) located downstream of the QTN. The results provide strong support for *IGF2* being the causative gene. The significant differences in *IGF2* mRNA expression between genotypes in skeletal and cardiac muscle and the lack of significant differences in fetal muscle and postnatal liver are consistent with our previous data showing clear phenotypic effects of the *IGF2* QTL on muscle growth and size of the heart but no effects on birth weight or weight of liver⁵. The results demonstrate that *IGF2* has an important role in regulating postnatal myogenesis. The higher expression in postnatal skeletal and cardiac muscle associated with the *Q* allele parallels the observation of a continued postnatal *IGF2* expression in mesodermal tissue of transgenic mice carrying a 5-kb deletion of *IGF2* encompassing DMR1 and the associated CpG island¹⁴.

Immunoreactive serum levels of IGF-II were determined by a radioimmunoassay, but no significant differences between genotypes were observed (Supplementary Fig. 5). This finding was expected because the major source of IGF-II in serum is generated from liver, where no difference in *IGF2* expression was detected (Fig. 3b). The results suggest that locally produced IGF-II in muscle determines the phenotype, as also indicated by the observation that there is no general overgrowth in pigs expressing the *Q* allele but rather a changed body composition.

There has been a strong selection for lean growth (high muscle mass and low fat content) in commercial pig populations over the past 50 yr. Therefore we investigated how this selection pressure has affected the allele frequency distribution of the *IGF2* QTL. The causative mutation was absent in a small sample of European and Asian wild boars and in breeds that have not been strongly selected for lean growth (Supplementary Table 1). In contrast, the causative mutation was found at high frequencies in several breeds that have been subjected to strong selection for lean growth. This confirms our prior assumption that *IGF2***Q* has experienced a selective sweep and has been spread between breeds by cross-breeding. The *IGF2***Q* mutation increases meat production, at the expense of fat, by 3–4%. The high frequency of *IGF2***Q* among major pig breeds implies that this mutation has had a large impact on pig production. European and Asian pigs were domesticated from different subspecies of the wild boar, and Asian germplasm was introgressed into European pig breeds during the eighteenth and nineteenth centuries¹⁶. The *IGF2***Q* mutation apparently occurred on an Asian chromosome as it shows a very close relationship to the haplotype carried by Chinese Meishan pigs. This explains the large genetic distance between *Q* and *q* haplotypes present in European domestic pigs (Fig. 1; see also Supplementary Fig. 1).

We have achieved an extraordinary resolution, down to a single nucleotide difference, in the genetic analysis of a QTL. This was

possible by exploiting a selective sweep of a favourable mutation in commercial pig populations. Determination of complete genome sequences from major farm animals in the near future will allow the exploitation of the full potential of farm animal genomics. High-density single-nucleotide polymorphism typing of domestic animals will allow identification of selective sweeps as documented here. Furthermore, re-sequencing of the entire genome using samples from different breeds, including wild ancestral species, will be very fruitful for studying genotype–phenotype relationships. This is well illustrated by the recent identification of the causative mutations for several interesting phenotypes in domestic animals^{17–23}. For instance, the *callipyge* mutation in sheep shares several features with the *IGF2* mutation in pigs; it affects body composition (muscle/fat content), shows a parent-of-origin effect and is a single-base substitution in a non-coding region²². In fact, farm animal genetics provides major advantages compared with human genetics, and in some aspects compared with model organisms, for genetic dissection of multifactorial traits². Farm animals are emerging as prime model organisms for understanding the genetic basis for multifactorial traits. □

Methods

Marker-assisted segregation analysis

QTL genotyping of the Pietrain/Large White, wild boar/Large White and Hampshire/Landrace crosses was performed as described⁸. Briefly, the likelihood of the pedigree data was computed under two hypotheses: H_0 , postulating that the corresponding boar was homozygous at the QTL (Q/Q or q/q), and H_1 , postulating that the boar was heterozygous Q/q . Likelihoods were computed using 'percentage lean meat' as phenotype, and assuming an allele substitution effect of 3.0%⁶. If the probability in favour of one of the hypotheses were superior or equal to 100:1, the most likely hypothesis was considered to be true. The Meishan/Large White cross consisted of 703 F₂ animals with data on back-fat depths. An interval analysis approach²⁴ with microsatellite markers spanning the *IGF2* region revealed no indication of an overall QTL effect, imprinted or not.

DNA sequencing

Animals homozygous for 13 of the haplotypes of interest were identified using flanking markers and pedigree information. A 28.6-kb segment was amplified from genomic DNA in seven long-range PCR products using the Expand Long Template PCR system (Roche Diagnostics GmbH). The same procedure was used to amplify the remaining M220 and LW197 haplotypes from two BAC clones isolated from a genomic library made from a Meishan/Large White F₁ individual²⁵. PCR products were purified using GeneClean (Polylab) and sequenced using the Big Dye Terminator Sequencing or dGTP Big Dye Terminator kits (Perkin Elmer). All primers are given in Supplementary Table 2. The sequence traces were assembled and analysed for DNA sequence polymorphism using Polyphred/Phrap/Consed²⁶.

Genotyping of *IGF2*-intron3-3072

Genotyping was done by pyrosequencing (Pyrosequencing AB). A 231-bp DNA fragment was PCR-amplified using Hot Star Taq DNA polymerase and Q-Solution (Qiagen) with the primers 18274F (5'-biotin-GGGCCGCGGCTTCGCTAG-3') and 18274R (5'-CGCACGCTTCCTCGCCACTG-3'). The sequencing primer (5'-CCCCACGCGCTCC CGCGCT-3') was designed on the reverse strand because of the palindrome located 5' of the QTN.

Bisulphite-based methylation analysis

Bisulphite sequencing was performed as described²⁷. A 300-bp fragment centred around intron3-3072 was amplified using a two-step PCR reaction with the following primers: PCR1-UP, 5'-TTGAGTGGGGATTGTTGAAGTTT-3'; PCR1-DN, 5'-ACCCACTTAT AATCTAAAAAATAATAATATATCTAA-3'; PCR2-UP, 5'-GGGGATTGTTGAAG TTTT-3'; PCR2-DN, 5'-CTTCTCCTACCCTAAATAA-3'. The amplified strand was chosen in order to differentiate the *Q* and *q* alleles. PCR products were cloned in the pCR2.1 vector (Invitrogen). Plasmid DNA was purified (Plasmid mini kit, Qiagen) and sequenced (Big Dye Terminator kit, Perkin Elmer). Inserts with identical sequences, that is having the same combination of C residues (whether part of a CpG dinucleotide or not) converted to U, were considered to derive from the same PCR product and were only considered once.

Electrophoretic mobility shift assays

DNA-binding proteins were extracted as described²⁸. EMSAs were performed with 40 fm ³²P-labelled, double-stranded oligonucleotide, 10 µg nuclear extract and 2 µg poly(dI-dC) in binding buffer (15 mM HEPES pH 7.65, 30.1 mM KCl, 2 mM MgCl₂, 2 mM spermidine, 0.1 mM EDTA, 0.63 mM dithiothreitol, 0.06% NP-40, 7.5% glycerol). For competition assays a 10-, 20-, 50- and 100-fold molar excess of cold double-stranded oligonucleotide were added. Reactions were incubated for 20 min on ice before ³²P-labelled, double-stranded oligonucleotide was added. Binding was allowed to proceed for 30 min at room temperature. DNA–protein complexes were resolved on a 5% native polyacrylamide gel

run in TBE $\times 0.5$ at room temperature for 2 h at 150 V. The following two (*Q/q*) 27-bp unmethylated oligonucleotides were used: 5'-GATCCTTCGGCTAGGCTC(A/G)CAGCG CGGGAGCGA-3'. A methylated *q* probe (*q**) was generated by incorporating a methylated cytosine at the mutated CpG site during oligonucleotide synthesis.

Transient transfection assay

The constructs contained 578 bp from *IGF2* intron 3 (nucleotides 2868–3446), followed by the *IGF2* P3 promoter (nucleotides –222 to +45 relative to the start of transcription)¹² and a luciferase reporter. C2C12 myoblast cells were grown to approximately 80% confluence. Cells were transiently co-transfected with the firefly luciferase reporter construct (4 μ g) and a Renilla luciferase control vector (pHRG-TK, Promega; 80 ng) using 10 μ g Lipofectamine 2000 (Invitrogen). Cells were incubated for 25 h before lysis in 100 μ l Triton lysis solution. Luciferase activities were measured using the Dual-Luciferase Reporter Assay System (Promega). The results are based on four triplicate experiments using two independent plasmid preparations for each construct. Statistical analysis was done with an analysis of variance.

Northern blot analysis and real-time RT-PCR

Total RNA was prepared using Trizol (Invitrogen) and treated with DNase I (Ambion). Products from the first-strand complementary DNA synthesis (Amersham Biosciences) were purified with QIAquick columns (Qiagen). Poly(A)⁺ RNA was then isolated using the Oligotex mRNA kit (Qiagen). Poly(A)⁺ mRNA (about 75 ng) from each sample was separated in a MOPS/formaldehyde agarose gel and transferred overnight to a Hybond-N⁺ nylon membrane (Amersham Biosciences). The membrane was hybridized in ExpressHyb hybridization solution (Clontech). The quantification of the transcripts was performed with a Phosphor Imager 425 (Molecular Dynamics). Real-time PCR was performed with an ABI PRISM 7700 instrument (Applied Biosystems). TaqMan probes and primers are given in Supplementary Table 3. PCRs were performed in triplicate using the Universal PCR Master Mix (Applied Biosystems). Messenger RNA was quantified using ten-point calibration curves established by dilution series of the cloned PCR products. Statistical evaluations were done with a two-sided Kruskal–Wallis rank-sum test.

Received 11 June; accepted 11 September 2003; doi:10.1038/nature02064.

- Mackay, T. F. C. Quantitative trait loci in *Drosophila*. *Nature Rev. Genet.* **2**, 11–21 (2001).
- Andersson, L. Genetic dissection of phenotypic diversity in farm animals. *Nature Rev. Genet.* **2**, 130–138 (2001).
- Glazier, A. M., Nadeau, J. H. & Aitman, T. J. Finding genes that underlie complex traits. *Science* **298**, 2345–2349 (2002).
- Darvasi, A. & Pisante-Shalom, A. Complexities in the genetic dissection of quantitative trait loci. *Trends Genet.* **18**, 489–491 (2002).
- Jeon, J.-T. *et al.* A paternally expressed QTL affecting skeletal and cardiac muscle mass in pigs maps to the *IGF2* locus. *Nature Genet.* **21**, 157–158 (1999).
- Nezer, C. *et al.* An imprinted QTL with major effect on muscle mass and fat deposition maps to the *IGF2* locus in pigs. *Nature Genet.* **21**, 155–156 (1999).
- King, M. C. & Wilson, A. C. Evolution at two levels in humans and chimpanzees. *Science* **188**, 107–116 (1975).
- Nezer, C. *et al.* Haplotype sharing refines the location of an imprinted QTL with major effect on muscle mass to a 250 Kb chromosome segment containing the porcine *IGF2* gene. *Genetics* **165**, 277–285 (2003).
- Florini, J. R., Ewton, D. Z. & McWade, F. J. IGFs, muscle growth, and myogenesis. *Diabetes Rev.* **3**, 73–92 (1995).
- Evans, G. J. *et al.* Identification of quantitative trait loci for production traits in commercial pig populations. *Genetics* **164**, 621–627 (2003).
- de Koning, D. J. *et al.* Genome-wide scan for body composition in pigs reveals important role of imprinting. *Proc. Natl Acad. Sci. USA* **97**, 7947–7950 (2000).
- Amarger, V. *et al.* Comparative sequence analysis of the *Insulin-IGF2-H19* gene cluster in pigs. *Mamm. Genome* **13**, 388–398 (2002).
- Greally, J. M., Guinness, M. E., McGrath, J. & Zemel, S. Matrix-attachment regions in the mouse chromosome 7F imprinted domain. *Mamm. Genome* **8**, 805–810 (1997).
- Constancia, M. *et al.* Deletion of a silencer element in *Igf2* results in loss of imprinting independent of *H19*. *Nature Genet.* **26**, 203–206 (2000).
- Eden, S. *et al.* An upstream repressor element plays a role in *Igf2* imprinting. *EMBO J.* **20**, 3518–3525 (2001).
- Giuffra, E. *et al.* The origin of the domestic pig: independent domestication and subsequent introgression. *Genetics* **154**, 1785–1791 (2000).
- Grobet, L. *et al.* A deletion in the bovine myostatin gene causes the double-muscling phenotype in cattle. *Nature Genet.* **17**, 71–74 (1999).
- Milan, D. *et al.* A mutation in *PRKAG3* associated with excess glycogen content in pig skeletal muscle. *Science* **288**, 1248–1251 (2000).
- Galloway, S. M. *et al.* Mutations in an oocyte-derived growth factor gene (*BMP15*) cause increased ovulation rate and infertility in a dosage-sensitive manner. *Nature Genet.* **25**, 279–283 (2000).
- Mulsant, P. *et al.* Mutation in bone morphogenetic protein receptor-IB is associated with increased ovulation rate in Booroola Merino ewes. *Proc. Natl Acad. Sci. USA* **98**, 5104–5109 (2001).
- Pailhoux, E. *et al.* A 11.7-kb deletion triggers intersexuality and polledness in goats. *Nature Genet.* **29**, 453–458 (2001).
- Freking, B. A. *et al.* Identification of the single base change causing the callipyge muscle hypertrophy phenotype, the only known example of polar overdominance in mammals. *Genome Res.* **12**, 1496–1506 (2002).
- Grisart, B. *et al.* Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine *DGAT1* gene with major effect on milk yield and composition. *Genome Res.* **12**, 222–231 (2002).
- Haley, C. S., Knott, S. A. & Elsen, J. M. Mapping quantitative trait loci in crosses between outbred lines using least squares. *Genetics* **136**, 1195–1207 (1994).

- Anderson, S. I., Lopez-Corralles, N. L., Gorick, B. & Archibald, A. L. A large fragment porcine genomic library resource in a BAC vector. *Mamm. Genome* **11**, 811–814 (2000).
- Nickerson, D., Tobe, V. O. & Taylor, S. L. PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescent-based resequencing. *Nucleic Acids Res.* **25**, 2745–2751 (1997).
- Engemann, S., El-Maarri, O., Hajkova, P., Oswald, J. & Walter, J. in *Methods in Molecular Biology* Vol. 181 (ed. Ward, A.) (Humana Press, Totowa, New Jersey, 2002).
- Andrews, N. C. & Faller, D. V. A rapid micropreparation technique for extraction of DNA-binding proteins from limiting numbers of mammalian cells. *Nucleic Acids Res.* **19**, 2499 (1991).
- Kashuk, C., Sengupta, S., Eichler, E. & Chakravarti, A. ViewGene: a graphical tool for polymorphism visualization and characterization. *Genome Res.* **12**, 333–338 (2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank C. Charlier and H. Ronne for discussions, M. Laita, B. McTeir, J. Pettersson, A.-C. Svensson and M. Köping-Höggård for technical assistance, and the Pig Improvement Company for providing DNA samples from Berkshire and Gloucester Old Spot pigs. This work was supported by the Belgian Ministère des Classes Moyennes et de l'Agriculture, the AgriFunGen program at the Swedish University of Agricultural Sciences, the Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning, Gentec, the UK Department for Environment, Food and Rural Affairs, the UK Pig Breeders Consortium, and the Biotechnology and Biological Sciences Research Council.

Competing interests statement The authors declare competing financial interests: details accompany the paper on www.nature.com/nature.

Correspondence and requests for materials should be addressed to L.A. (Leif.Andersson@imbim.uu.se) or M.G. (michel.georges@ulg.ac.be). The sequence data reported in this paper have been deposited in GenBank under accession numbers AY242098–AY242112.

Identification of the haematopoietic stem cell niche and control of the niche size

Ji Wang Zhang¹, Chao Niu¹, Ling Ye², Haiyang Huang², Xi He¹, Wei-Gang Tong¹, Jason Ross¹, Jeff Haug¹, Teri Johnson¹, Jian Q. Feng², Stephen Harris², Leanne M. Wiedemann^{1,4}, Yuji Mishina³ & Linheng Li^{1,4}

¹Stowers Institute for Medical Research, Kansas City, Missouri 64110, USA
²Department of Oral Biology, School of Dentistry, University of Missouri–Kansas City, 650 East 25th Street, Kansas City, Missouri 64108, USA
³Laboratory of Reproductive and Developmental Toxicology, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 27709, USA
⁴Department of Pathology and Laboratory Medicine, Kansas University Medical Center, Kansas City, Kansas 66160, USA

Haematopoietic stem cells (HSCs) are a subset of bone marrow cells that are capable of self-renewal and of forming all types of blood cells (multi-potential)¹. However, the HSC 'niche'—the *in vivo* regulatory microenvironment where HSCs reside—and the mechanisms involved in controlling the number of adult HSCs remain largely unknown. The bone morphogenetic protein (BMP) signal has an essential role in inducing haematopoietic tissue during embryogenesis^{2,3}. We investigated the roles of the BMP signalling pathway in regulating adult HSC development *in vivo* by analysing mutant mice with conditional inactivation of BMP receptor type IA (BMPRIA). Here we show that an increase in the number of spindle-shaped N-cadherin⁺CD45⁺ osteoblastic (SNO) cells correlates with an increase in the number of HSCs. The long-term HSCs are found attached to SNO cells. Two adherens junction molecules, N-cadherin and β -catenin, are asymmetrically localized between the SNO cells and the long-term HSCs. We conclude that SNO cells lining the bone surface function as a key component of the niche to support HSCs, and that BMP signalling through BMPRIA controls the number of HSCs by regulating niche size.

We first analysed the expression patterns of *Bmpr1a* (*Alk3*) and *Bmpr1b* (*Alk6*) in bone marrow and the surrounding bone tissue. BMPRIA was found in bone marrow cells, including most of the haematopoietic lineages apart from the HSC population, and most osteoblastic cells (Fig. 1a–c). By contrast, we were unable to detect expression of *Bmpr1b* in HSCs and the other haematopoietic lineages that we examined (Fig. 1d).

To block the BMP signal, we inactivated BMPRIA by crossing *Bmpr1a^{fx/fx}* and *Bmpr1a^{fx/+}* mouse lines⁴ with a PolyI:C-inducible *Mx1-Cre* mouse line⁵ and assaying the mutant mice. Multiple injections of PolyI:C were required for efficient deletion of *Bmpr1a* (see Supplementary Information, data 1)

Mice with different injection schedules (see Methods) were investigated using flow cytometric assays to analyse the HSC population. The percentage of lineage-negative (*Lin*[−]) *Sca-1*⁺*c-Kit*⁺ (HSC) cells was increased 2-fold in the *Mx1-Cre*⁺*Bmpr1a^{fx/fx}* and 2.4-fold in the *Mx1-Cre*⁺*Bmpr1a^{fx/+}* mutant mice compared with the littermate controls (Fig. 2a). Although the total bone marrow cell number per femur was reduced in both of the *Bmpr1a* mutant lines owing to reduced cavity room, as seen in Fig. 3b, the absolute number of HSCs per femur was still increased 1.5–2-fold (Fig. 2b). As no significant difference between *Mx1-Cre*⁺*Bmpr1a^{fx/fx}* and *Mx1-Cre*⁺*Bmpr1a^{fx/+}* mutant mice was observed, we focused on the *Mx1-Cre*⁺*Bmpr1a^{fx/fx}* mice in the following studies.

The HSC population is a heterogeneous mixture, including long-term (LT) and short-term (ST) HSCs⁶. Therefore, we injected the mice with 5-bromodeoxyuridine (BrdU) to label cycling cells. Three hours after labelling, we analysed the HSC population to distinguish ST-HSCs from LT-HSCs according to differences in their cell-cycle state. The percentage of BrdU-negative cells in the HSC population (including the quiescent LT-HSCs) increased by an average of 2.4 times in the *Bmpr1a* mutant mice compared with littermate controls. The percentage of BrdU-positive cells in the HSC population

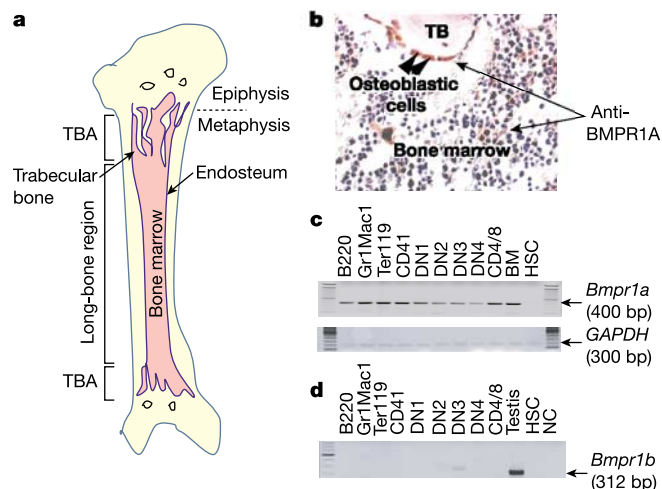


Figure 1 Expression patterns of BMPRIA and BMPRIb in bone and bone marrow. **a**, Schematic of bone and bone marrow structures. TBA, trabecular bone area. **b**, Expression of BMPRIA in osteoblastic cells and bone marrow cells. Bone/bone marrow sections were immunohistochemically stained using anti-BMPRIA serum. TB, trabecular bone. **c**, Detection of *Bmpr1a* gene expression in HSCs and lineage cells by RT-PCR (polymerase chain reaction with reverse transcription) assay. GAPDH, glyceraldehyde-3-phosphate; B220, B-cell marker; Gr1Mac1, myeloid lineage markers; Ter119, erythroid lineage marker; CD41, megakaryocyte marker; DN1, double negative for CD4 and CD8; DN2, CD44⁺CD25[−]; DN3, CD44[−]CD25⁺; DN4, CD44[−]CD25[−]. **d**, Detection of *Bmpr1b* gene expression in HSCs and lineage cells by RT-PCR assay. NC, negative control.

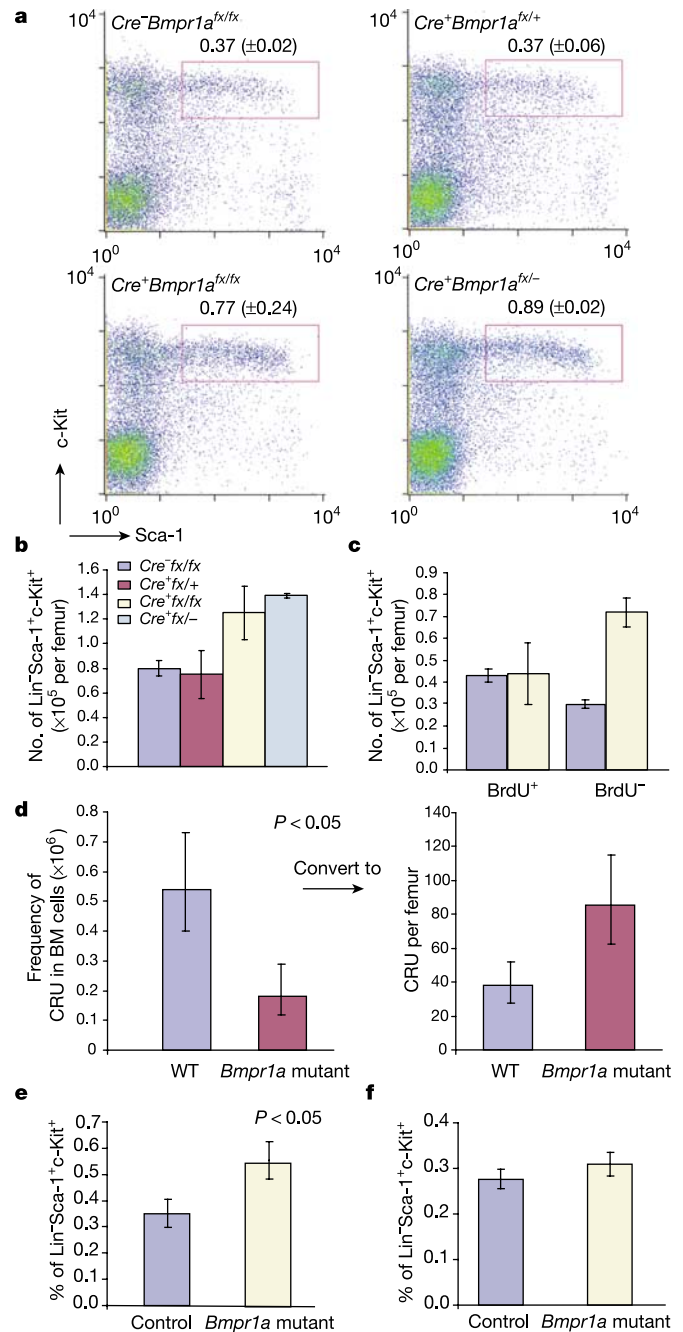


Figure 2 Analyses of the HSC population. For raw data, see Supplementary Tables 1–4. **a**, Typical example of flow assay on the HSC population. Figure in the gate is the percentage of HSCs in the BM-MNC fraction. **b**, Comparison of absolute number of HSCs per femur in wild-type and *Bmpr1a* mutant mice. Owing to the reduced bone cavity in *Bmpr1a* mutant mice, the total bone marrow cell number is also reduced in mutant mice (an average of 1.58×10^7 versus 2.1×10^7 in wild-type control mice). The absolute HSC number is equal to the percentage of HSCs multiplied by the total number of BM-MNCs. **c**, Comparison of the number of LT-HSCs and ST-HSCs in the *Bmpr1a* mutant and control mice. **d**, CRU assay to compare the number of functional HSCs per femur between the *Bmpr1a* mutant and wild-type (WT) control mice. The number of CRUs per femur (right panel) equals the frequency of CRUs (left panel) multiplied by the total number of BM-MNCs as indicated in **b**. **e–f**, Bar graphs of the percentage of HSCs in recipient mice. Wild-type donor (Ly5.1) HSCs (2×10^3) were transplanted into *Bmpr1a* mutant recipients or littermate controls (Ly5.2) (**e**). Bone marrow cells (2×10^6) from controls and mutants (Ly5.2) were transplanted into wild-type recipients (Ly5.1) (**f**).

(including the cycling ST-HSCs) was similar to that in the littermate controls (Fig. 2c). Therefore, direct comparison of the entire HSC population provides an underestimated but reliable representation (Fig. 2b; 1.6-fold) of the difference in LT-HSC numbers (Fig. 2c; 2.4-fold) between the mutant and control animals.

These experiments demonstrated an increase in the LT-HSC population. However, the gold-standard test for functional stem cells is the competitive repopulation unit (CRU) assay⁷, in which a series of diluted donor-derived bone marrow mononuclear cells (BM-MNCs) is transplanted into different groups of sublethally irradiated recipient mice. Three months after transplantation, peripheral blood is examined to monitor the engraftment of donor cells in the recipient mice in order to determine the lowest number of bone marrow cells required for the reconstitution of haematopoiesis. Using this assay, we confirmed that the functional stem cell number increased 2.2-fold in the mutant mice compared

with controls (Fig. 2d and Supplementary Table 4). The results from both immunophenotypical and CRU assays indicate that the population of LT-HSCs is expanded in the *Bmpr1a* mutant mice.

There are several mechanisms that can lead to changes in the HSC number: (1) an intrinsic change in stem cells that either promotes self-renewal or blocks apoptosis; (2) an internal defect in progenitors that inhibits differentiation, leading to an accumulation of stem cells; or (3) an external influence from the HSC microenvironment. The finding that *Bmpr1a* is not expressed in HSCs (Fig. 1c) does not support the argument that an intrinsic defect in HSCs was the cause of a change in the number of HSCs in the *Bmpr1a* mutant mice. We then analysed myeloid and lymphoid lineages, common myeloid progenitor and common lymphoid progenitor subsets^{8,9}. The results revealed no significant phenotypical change in any of the subsets analysed (Supplementary Information, data 2a, b). This was consistent with results obtained from *in vitro* colony-forming unit

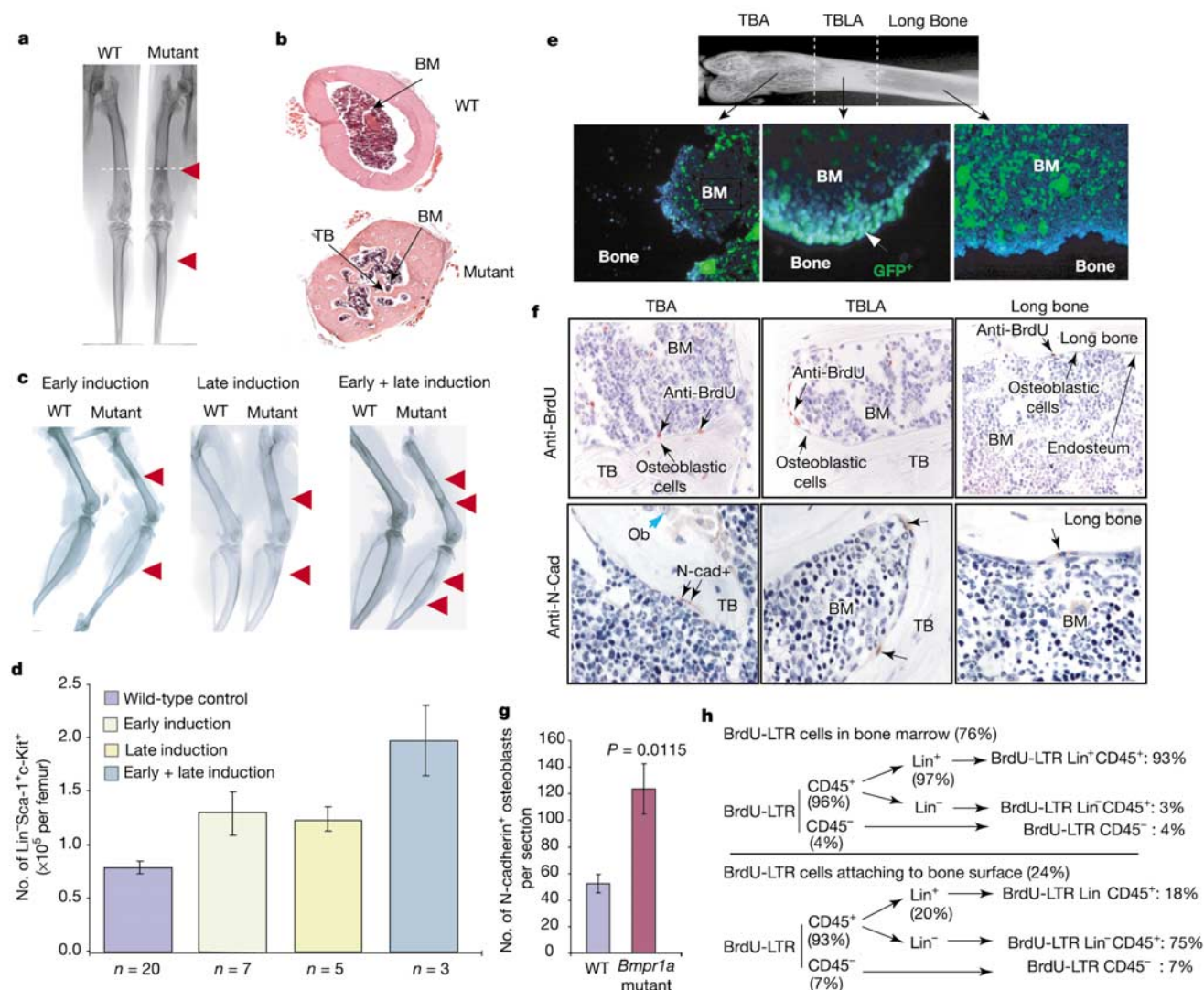


Figure 3 Analyses of bone structure and HSC number. **a**, **b**, X-ray image and histology of the bone structures of ectopically formed TBLA (red arrow). BM, bone marrow. **c**, X-ray images of femur and tibia from *Bmpr1a* mutant and control mice. **d**, Comparison of the absolute number of HSCs per femur in *Bmpr1a* mutant and wild-type control mice. **e**, Inactivation of *Bmpr1a* on the surface of ectopically formed TBLA in the triple-genotype mice as shown by GFP expression, counterstained with 4,6-diamidino-2-phenylindole

(DAPI). **f**, BrdU-LTR cells (upper panels, black arrow), SNO cells (lower panels, black arrow) and matrix-forming osteoblasts (blue arrow). **g**, SNO cell number in control and *Bmpr1a* mutant mice. (The average total number of the SNO cells per section is given in Supplementary Table 6.) **h**, Summary of the distribution of three types of BrdU-LTR cell (see text and Supplementary Table 7).

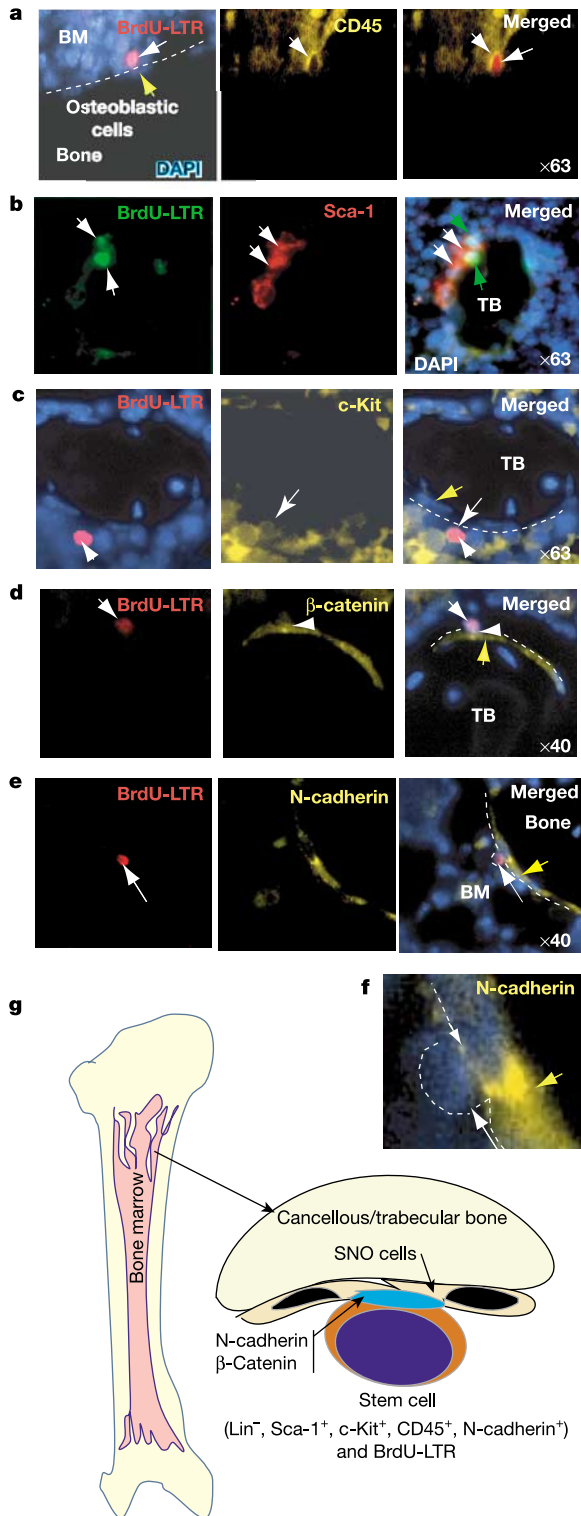


Figure 4 Co-staining of BrdU-LTR cells with HSC markers, counterstained with DAPI. **a–e**, HSC markers as indicated. The lens magnification used for the microscopic photograph is indicated. Most BrdU-LTR cells attached to the trabecular bone surface are co-stained with Sca-1 (74%), c-Kit (72%) or a pan-haematopoietic marker, CD45 (92%). Yellow arrow indicates SNO cells. **f**, A blow-up of part **e**. **g**, Model illustrating the haematopoietic stem cell niche. The SNO cells, which are located mainly on the surface of cancellous/trabecular bone, are a key component of the resident niche for HSCs.

(CFU) assays, which showed that the progenitor number of different lineages was similar in the *Bmpr1a* mutant and littermate control mice (Supplementary Information, data 3). These results ruled out the possibility of an accumulation of HSCs resulting from a block in progenitor cell differentiation.

Direct evidence of an external influence causing the change in HSC number came from reciprocal bone marrow/HSC transplantation experiments. HSCs isolated from wild-type Ly5.1 mice were transplanted into lethally irradiated *Bmpr1a* mutant and littermate control mice (both Ly5.2 genotype). The result of a flow cytometric assay three months after transplantation revealed that the percentage of donor-derived HSCs in the *Bmpr1a* mutant mice was, on average, 1.6 times higher than in the wild-type control recipient mice (Fig. 2e). A reciprocal assay was also carried out by transplanting BM-MNCs derived from the *Bmpr1a* mutant or wild-type control mice into lethally irradiated wild-type Ly5.1 mice. These flow cytometric results showed that the percentage of donor-derived HSCs in both groups of recipient mice was similar, regardless of the origin of BM-MNCs (Fig. 2f). We conclude that the change in the microenvironment led to an increase in HSC number in the *Bmpr1a* mutant mice.

Bone marrow stromal cells are derived from mesenchymal stem cells, including fibroblasts, adipocytes, endothelial cells and osteoblasts¹⁰. Previous studies of the roles of stromal cells in supporting HSCs have been based mainly on *in vitro* culture, and each of these stromal cells has been suggested to be capable of supporting haematopoietic stem/progenitor cells *in vitro*^{11,12}. However, none of these results has been confirmed *in vivo*.

We observed obvious abnormal bone formation in the *Bmpr1a* mutant mice. Our X-ray and histological analyses revealed that an ectopic formation of trabecular-bone-like area (TBLA) occurred in the long-bone region of the mutant mice (Fig. 3a, b). The location of the ectopically formed TBLAs varied depending on the PolyI:C injection time: TBLAs were seen distal to the knee in the early-induced group, but were proximal to the knee in the late-induced group, and at both sites in the combined-injection group (Fig. 3c). This raised the possibility of a change in osteoblasts or osteoclasts affecting the HSC microenvironment. Recent *in vitro* evidence shows that an osteoblastic cell line can expand the number of HSCs 2–4-fold¹². In addition, osteoblasts, when co-transplanted with HSCs, can also increase the engraftment rate¹³. These observations suggest that osteoblasts have a role in supporting HSCs.

We next asked whether the ectopically formed TBLA was responsible for the increased HSC number in the mutant mice. As described above (Fig. 2b), the number of HSCs per femur in the mutant animals was increased 1.67-fold on average in both early- and late-induced mutant mice. In addition, we found that the number of HSCs per femur increased 2.5-fold on average in the combined early- and late-induced mice, in which two regions of TBLA were observed (Fig. 3d). This shows that an increase in the number of HSCs correlates with an increase in the number of ectopically formed TBLAs.

We investigated the mechanism that leads to the ectopic formation of TBLA when the BMP signal is blocked. BMPRIA has been shown to inhibit osteoblastic lineage commitment from mesenchymal progenitors *in vitro*¹⁴. We ruled out the possibility of an increase in the number of mesenchymal progenitors in the mutants using a CFU-fibroblast assay¹⁰ (Supplementary Table 5). Osteoblastic lineage commitment from mesenchymal progenitors is not increased, because the ectopic formation of TBLA is regional rather than evenly distributed. This suggests that regional inactivation of *Bmpr1a* leads to the ectopic formation of TBLA. To confirm this, we generated a triple-genotype mouse line bearing *Mx1-Cre*, *Bmpr1a*^{flx/flx} and *Z/EG* alleles. Cre-induced green fluorescent protein (GFP) expression reflects successful targeting of *Bmpr1a* (Fig. 3e). Analysis of bone sections derived from the triple-genotype mice revealed that only the surface of ectopically formed TBLA was GFP

positive, indicating that regional deletion of *Bmpr1a* occurs exclusively in the cells that line the TBLA surface. This was further supported by a 3-fold increase in osteoblast number, a 2-fold increase in the rate of bone formation, and a 10-fold increase in bone volume in the TBLA of *Bmpr1a* mutants compared with the long-bone region in control littermates (Supplementary Information, data 4, and Supplementary Table 8); however, there was no apparent change in osteoclasts (cells that reabsorb bone) measured by tartrate-resistant acid phosphatase staining (data not shown). The ectopic TBLA might be formed by over-proliferation or abnormal differentiation of an osteoblast progenitor. In addition, inactivation of BMPRIA might make the osteoblasts less sensitive to apoptotic signal, as a BMP signal is known to induce osteoblast apoptosis¹⁵.

We observed a correlation between the number of LT-HSCs and the increase in TBLA. Are the LT-HSCs in the TBLA enriched? The limited number of LT-HSCs in bone marrow and the lack of a unique marker make their visualization difficult. As LT-HSCs are quiescent or slow cycling, they are able to retain labelled nucleotides for a relatively long period and can be identified as BrdU-LTR (long-term retaining) cells¹⁶ (see Methods). We found that 76% of the BrdU-LTR cells were located within the bone marrow cavity, whereas 24% were attached to the bone surface. Most BrdU-LTR cells attached to the bone surface were located in cancellous/trabecular bone area (including epiphysis and metaphysis), whereas the rest were dispersed along the endosteal surface of long bone (Fig. 3f, upper panel). This is consistent with previous observations in which HSCs were found to be close to the endosteal surface of long bone¹⁷ or homing to the bone surface of epiphysis¹⁸. Importantly, a significantly increased number of BrdU-LTR cells was found in the ectopically formed TBLA compared with the long-bone region (Fig. 3f). There are different types of BrdU-LTR cell, including Lin⁻CD45⁺ (enriched with LT-HSCs), Lin⁺CD45⁺ (enriched with memory T or B cells) and Lin⁻CD45⁻ (enriched with mesenchymal stem cells) (see Fig. 3h). We estimate that the frequency of the enriched LT-HSCs in control animals (BrdU-LTR Lin⁻CD45⁺ cells) is 0.01%, close to the frequency of LT-HSCs (0.007%) on the basis of functional studies¹⁹. These cells are located mainly on the bone surface, particularly the surface of the cancellous/trabecular bone, and most of these cells are co-stained with other HSC markers: CD45, Sca-1 and c-Kit (Fig. 4a–c). These observations support the conclusion that the ectopically formed TBLA is responsible for the increase in the number of HSCs.

The increased number of osteoblasts on the surface of ectopically formed TBLA correlates with the increase in HSC number. The LT-HSCs appear to be attached to cells with an early osteoblastic character (the mononuclear spindle-shaped cells lining the bone surface; Fig. 3f, upper panel). As N-cadherin is expressed in both early and late osteoblastic cells²⁰, we used this biological marker to confirm this observation. This marker stains two osteoblastic cell types: a small subset of spindle-shaped osteoblasts (osteoblastic lining cells) and most of the larger, oval-shaped, matrix-forming osteoblasts (Fig. 3f, lower panel). Histological analysis of the distribution of the spindle-shaped N-cadherin⁺ osteoblastic (SNO) cells revealed that these cells were enriched on the surface of cancellous/trabecular bone, including the vesicle area in epiphysis, and sporadically dispersed along the endosteal surface of long bone (Fig. 3f, lower panel). This distribution pattern was similar to that of LT-HSCs (Fig. 3f, upper panel). Indeed, the LT-HSCs are attached only to the SNO cells that line the bone surface, (Fig. 4e). We also counted the number of SNO cells (Supplementary Table 6). The result showed that an increase in the number of these cells (2.3-fold; Fig. 3g) correlates to a high degree with the increase in LT-HSC number (2.2-fold; Fig. 2d) in the *Bmpr1a* mutant mice. Taken together, these observations indicate that the SNO cells have an important role in supporting LT-HSCs. This conclusion is also

supported by evidence from studies of conditional ablation of osteoblasts in Col2.3 Δ tk (thymidine kinase) transgenic mice, which showed that osteoblasts were necessary for the maintenance of haematopoiesis and where a loss of osteoblasts led to loss of haematopoietic cells²¹.

As the functions of the niche include adhesive interaction between stem cells and the niche²², we asked whether the SNO cells provide an adhesive attachment for HSCs. In *Drosophila*, two important junction-related adherens molecules, E-cadherin and β -catenin, are essential for the maintenance of ovarian somatic stem cells, as evidenced by the observation that a loss of E-cadherin leads to a loss of somatic stem cells²³. E-cadherin was not expressed in either the osteoblasts or the LT-HSCs (although it was expressed in many bone marrow cells; data not shown). However, we found that N-cadherin was asymmetrically localized to the cell surface of LT-HSCs adjacent to the SNO cells (Fig. 4e, f). Using flow cytometric assay, we confirmed that N-cadherin is expressed in a subpopulation (10%) of murine adult HSCs (Lin⁻Sca-1⁺c-Kit⁺) (Supplementary Information, data 5). It will be important to demonstrate an *in vivo* functional role of the N-cadherin⁺ HSCs. In addition, β -catenin, which interacts with and forms an adherens complex with N-cadherin²⁴, is also found to be asymmetrically localized between the SNO cells and the LT-HSCs (Fig. 4d).

The niche hypothesis was first proposed by Schofield²⁵ in 1978, and is supported by the co-culture of HSCs with marrow stromal cells²⁶. The stem cell niche has been described in the *Drosophila* ovary²⁷, but the HSC niche, owing to complicated anatomical architecture, has remained an enigma. We show that, in bone and bone marrow, the cancellous/trabecular bone area is the primary site for HSCs. The SNO cells located on the bone surface are a key component of the niche, supporting LT-HSCs (Fig. 4g). SNO cells might support HSCs through a specific adhesive interaction between N-cadherin and β -catenin. Consistent with this, in *coll-1*-PPR (parathyroid hormone (PTH)/PTHrP receptor) transgenic mice, the osteoblastic lineage has been defined as a key participant in the regulation of HSC numbers²⁸. The niche size must be tightly regulated *in vivo* to maintain HSCs and normal homeostasis^{25,29}. We provide *in vivo* evidence to show that a change in the niche size affects the number of stem cells. BMP signalling through BMPRIA is an important component of this regulatory system. □

Methods

Inducible Cre expression

PolyI:C (250 μ g per mouse) was injected intraperitoneally. For the early-induced group, PolyI:C was injected on the 3rd, 5th and 7th day after birth, whereas the late-induced group was treated with PolyI:C on the 21st, 23rd and 25th day after birth. A third double-induced group combined both injection schedules.

Flow cytometric assay and CFU culture

Isolation and preparation of bone marrow, thymus, spleen and peripheral blood cells, and the method for subsequent flow cytometric assays have been described³⁰. For HSC analysis, BM-MNCs were stained with fluorescein isothiocyanate (FITC)-lineage markers (CD4, CD8, CD3, B220, IgM, Mac-1, Gr-1 and Ter-119) to eliminate Lin⁺ cells, but positive for APC-c-Kit and phycoerythrin (PE)-Sca-1. For lineage analyses, see Supplementary Information, data 1a, b. CFU assay was performed according to the manufacturer's recommendations (StemCell Technologies).

CRU assay

A series of diluted BM-MNCs (1×10^4 , 3×10^4 , 1×10^5 and 3×10^5) from either *Bmpr1a* mutant or wild-type control were transplanted into several groups of sublethally (500 rad) irradiated Ly5.1 recipient female mice. Three months after transplantation, peripheral blood was analysed using myeloid and lymphoid lineage markers: Mac-1, Gr-1, B220, CD3 and Ly5.2. An engraftment rate of >1% (the basal level was defined by transplantation of bone marrow derived from Ly5.1 mice) was scored as positive. The data shown in Fig. 2f are based on two independent experiments. The CRU frequencies were determined using L-Cal software (StemCell Technologies), which uses Poisson statistics and the method of maximum likelihood to the proportion of negative recipients, using limiting-dilution analysis (see Supplementary Table 4). We found that the CRU frequency was lower than that reported in studies where lethal doses of irradiation have been used⁷.

Immunohistochemistry and BrdU-LTR assay

The procedures for bone and bone marrow section preparation, immunostaining conditions and antibodies are described in Supplementary Methods. The procedure for BrdU pulse labeling, LTR and subsequent detection has been reported¹⁶. The mice were fed BrdU (0.8 mg ml⁻¹ in water) for 10 days, during which time 40% of LT-HSCs would divide at least once³¹. Seventy days after BrdU labelling, sections were stained with anti-BrdU antibody.

N-cadherin⁺ cell count

For quantitative analysis of N-cadherin⁺ cells, the sections were developed with AEC after being incubated with rabbit anti-N-cadherin antibody for 1 h and horseradish peroxidase (HRP)-conjugated goat anti-rabbit second antibody for 1 h. Three people counted the SNO cells in these sections, blind to the source of the sections.

X-ray image

High-resolution X-rays (Faxitron MX-20) of bone and bone histomorphometry (OsteoMetrics, Inc.) were performed at the University of Missouri–Kansas City School of Dentistry.

Received 12 May; accepted 12 August 2003; doi:10.1038/nature02041.

- Spangrude, G. J., Heimfeld, S. & Weissman, I. L. Purification and characterization of mouse hematopoietic stem cells. *Science* **241**, 58–62 (1988).
- Maeno, M. *et al.* The role of BMP-4 and GATA-2 in the induction and differentiation of hematopoietic mesoderm in *Xenopus laevis*. *Blood* **88**, 1965–1972 (1996).
- Davidson, A. J. & Zon, L. I. Turning mesoderm into blood: the formation of hematopoietic stem cells during embryogenesis. *Curr. Top. Dev. Biol.* **50**, 45–60 (2000).
- Mishina, Y., Hanks, M. C., Miura, S., Tallquist, M. D. & Behringer, R. R. Generation of *Bmpr/Alk3* conditional knockout mice. *Genesis* **32**, 69–72 (2002).
- Kuhn, R., Schwenk, F., Aguet, M. & Rajewsky, K. Inducible gene targeting in mice. *Science* **269**, 1427–1429 (1995).
- Morrison, S. J. & Weissman, I. L. The long-term repopulating subset of hematopoietic stem cells is deterministic and isolatable by phenotype. *Immunity* **1**, 661–673 (1994).
- Szilyassy, S., Nicolini, F. E., Eaves, C. J. & Miller, C. L. in *Hematopoietic Stem Cell Protocols* (eds Klug, C. A. & Jordan, C. T.) 167–187 (Humana Press, Totowa, New Jersey, 2002).
- Kondo, M., Weissman, I. L. & Akashi, K. Identification of clonogenic common lymphoid progenitors in mouse bone marrow. *Cell* **91**, 661–672 (1997).
- Akashi, K., Traver, D., Miyamoto, T. & Weissman, I. L. A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature* **404**, 193–197 (2000).
- Simmons, P., Gronthos, S. & Zannettino, A. C. in *Hematopoiesis: A Developmental Approach* (ed. Zon, L. I.) 718–726 (Oxford Univ. Press, New York, 2001).
- Charbord, P. The hematopoietic stem cell and the stromal microenvironment. *Therapie* **56**, 383–384 (2001).
- Taichman, R. S., Reilly, M. J. & Emerson, S. G. The hematopoietic microenvironment: osteoblasts and the hematopoietic microenvironment. *Hematology* **4**, 421–426 (2000).
- El-Badri, N. S., Wang, B. Y. & Cherry Good, R. A. Osteoblasts promote engraftment of allogeneic hematopoietic stem cells. *Exp. Hematol.* **26**, 110–116 (1998).
- Chen, D. *et al.* Differential roles for bone morphogenetic protein (BMP) receptor type IB and IA in differentiation and specification of mesenchymal precursor cells to osteoblast and adipocyte lineages. *J. Cell Biol.* **142**, 295–305 (1998).
- Hay, E., Lemmonier, J., Fromiguet, O. & Marie, P. J. Bone morphogenetic protein-2 promotes osteoblast apoptosis through a Smad-independent, protein kinase C-dependent signaling pathway. *J. Biol. Chem.* **276**, 29028–29036 (2001).
- Taylor, G., Lehrer, M. S., Jensen, P. J., Sun, T. T. & Lavker, R. M. Involvement of follicular stem cells in forming not only the follicle but also the epidermis. *Cell* **102**, 451–461 (2000).
- Nilsson, S. K., Johnston, H. M. & Coverdale, J. A. Spatial localization of transplanted hematopoietic stem cells: inferences for the localization of stem cell niches. *Blood* **97**, 2293–2299 (2001).
- Askenasy, N. & Farkas, D. L. Optical imaging of PKH-labeled hematopoietic cells in recipient bone marrow *in vivo*. *Stem Cells* **20**, 501–513 (2002).
- Lagasse, E., Shizuru, J. A., Uchida, N., Tsukamoto, A. & Weissman, I. L. Toward regenerative medicine. *Immunity* **14**, 425–436 (2001).
- Hay, E. *et al.* N- and E-cadherin mediate early human calvaria osteoblast differentiation promoted by bone morphogenetic protein-2. *J. Cell. Physiol.* **183**, 117–128 (2000).
- Visnjic, D. *et al.* Conditional ablation of the osteoblast lineage in Col2.3 α tk transgenic mice. *J. Bone Miner. Res.* **16**, 2222–2231 (2001).
- Spradling, A., Drummond-Barbosa, D. & Kai, T. Stem cells find their niche. *Nature* **414**, 98–104 (2001).
- Song, X. & Xie, T. DE-cadherin-mediated cell adhesion is essential for maintaining somatic stem cells in the *Drosophila* ovary. *Proc. Natl Acad. Sci. USA* **99**, 14813–14818 (2002).
- Puch, S. *et al.* N-cadherin is developmentally regulated and functionally involved in early hematopoietic cell differentiation. *J. Cell Sci.* **114**, 1567–1577 (2001).
- Schofield, R. The relationship between the spleen colony-forming cell and the hematopoietic stem cell. A hypothesis. *Blood Cells* **4**, 7–25 (1978).
- Dexter, T. M., Allen, T. D. & Lajtha, L. G. Conditions controlling the proliferation of hematopoietic stem cells *in vitro*. *J. Cell. Physiol.* **91**, 335–344 (1977).
- Xie, T. & Spradling, A. C. A niche maintaining germ line stem cells in the *Drosophila* ovary. *Science* **290**, 328–330 (2000).
- Calvi, L. M. *et al.* Osteoblastic cells regulate the hematopoietic stem cell niche. *Nature* **425**, 841–846 (2003).
- Weissman, I. L. Developmental switches in the immune system. *Cell* **76**, 207–218 (1994).
- Akashi, K. *et al.* Transcriptional accessibility for genes of multiple tissues and hematopoietic lineages is hierarchically controlled during early hematopoiesis. *Blood* **101**, 383–389 (2003).
- Cheshier, S. H., Morrison, S. J., Liao, X. & Weissman, I. L. *In vivo* proliferation and cell cycle kinetics of long-term self-renewing hematopoietic stem cells. *Proc. Natl Acad. Sci. USA* **96**, 3120–3125 (1999).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank P. Dijke for providing anti-BMPRIA anti-serum, and C. G. Lobe for Z/EG reporter mice. We thank R. Krumlauf and X. Liao for scientific discussion. We thank H. Lin for critically reviewing the manuscript, and L. Bonewald for training in bone histomorphometry. We are grateful to D. di Natale and T. Langner for assistance in manuscript editing. We are grateful to S. Morris and D. Stark for imaging assistance, W. Walker and her co-workers for animal care, and C. Sonnenbrot and his co-workers for assistance in medium preparation. This work was supported by the Stowers Institute for Medical Research.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to L.L. (lil@stowers-institute.org).

Osteoblastic cells regulate the haematopoietic stem cell niche

L. M. Calvi^{1*}, G. B. Adams^{3*}, K. W. Weibrecht³, J. M. Weber¹, D. P. Olson³, M. C. Knight⁴, R. P. Martin³, E. Schipani⁴, P. Divieti⁴, F. R. Bringhurst⁴, L. A. Milner², H. M. Kronenberg⁴ & D. T. Scadden³

¹Endocrine Unit, Department of Medicine, and ²Department of Pediatrics, Center for Human Genetics and Molecular Pediatric Disease, University of Rochester School of Medicine, Rochester, New York 14642, USA

³Center for Regenerative Medicine and Technology, Partners AIDS Research Center, MGH Cancer Center, and ⁴Endocrine Unit, Massachusetts General Hospital, Harvard Medical School, Boston 02114, Massachusetts, USA

* These authors contributed equally to this work

Stem cell fate is influenced by specialized microenvironments that remain poorly defined in mammals^{1–3}. To explore the possibility that haematopoietic stem cells derive regulatory information from bone, accounting for the localization of haematopoiesis in bone marrow, we assessed mice that were genetically altered to produce osteoblast-specific, activated PTH/PTHrP receptors (PPRs)⁴. Here we show that PPR-stimulated osteoblastic cells that are increased in number produce high levels of the Notch ligand jagged 1 and support an increase in the number of haematopoietic stem cells with evidence of Notch1 activation *in vivo*. Furthermore, ligand-dependent activation of PPR with parathyroid hormone (PTH) increased the number of osteoblasts in stromal cultures, and augmented *ex vivo* primitive haematopoietic cell growth that was abrogated by γ -secretase inhibition of Notch activation. An increase in the number of stem cells was observed in wild-type animals after PTH injection, and survival after bone marrow transplantation was markedly improved. Therefore, osteoblastic cells are a regulatory component of the haematopoietic stem cell niche *in vivo* that influences stem cell function through Notch activation. Niche constituent cells or signalling pathways provide pharmacological targets with therapeutic potential for stem-cell-based therapies.

Mammalian bone marrow architecture involves haematopoietic stem cells (HSCs) in close proximity to the endosteal surfaces^{5,6}, with more differentiated cells arranged in a loosely graduated fashion as the central longitudinal axis of the bone is approached^{5,7,8}. This nonrandom organization of the marrow suggests a possible relationship between HSCs and osteoblasts—osteogenic cells lining the endosteal surface. Osteoblasts produce haematopoietic growth factors^{9–11} and are activated by parathyroid hormone (PTH) or the locally produced PTH-related protein (PTHrP), through the PTH/PTHrP receptor (PPR). We tested whether osteoblasts contribute to the unique microenvironment of the bone marrow *in vivo* using a constitutively active PPR (col1-caPPR) under the control of the

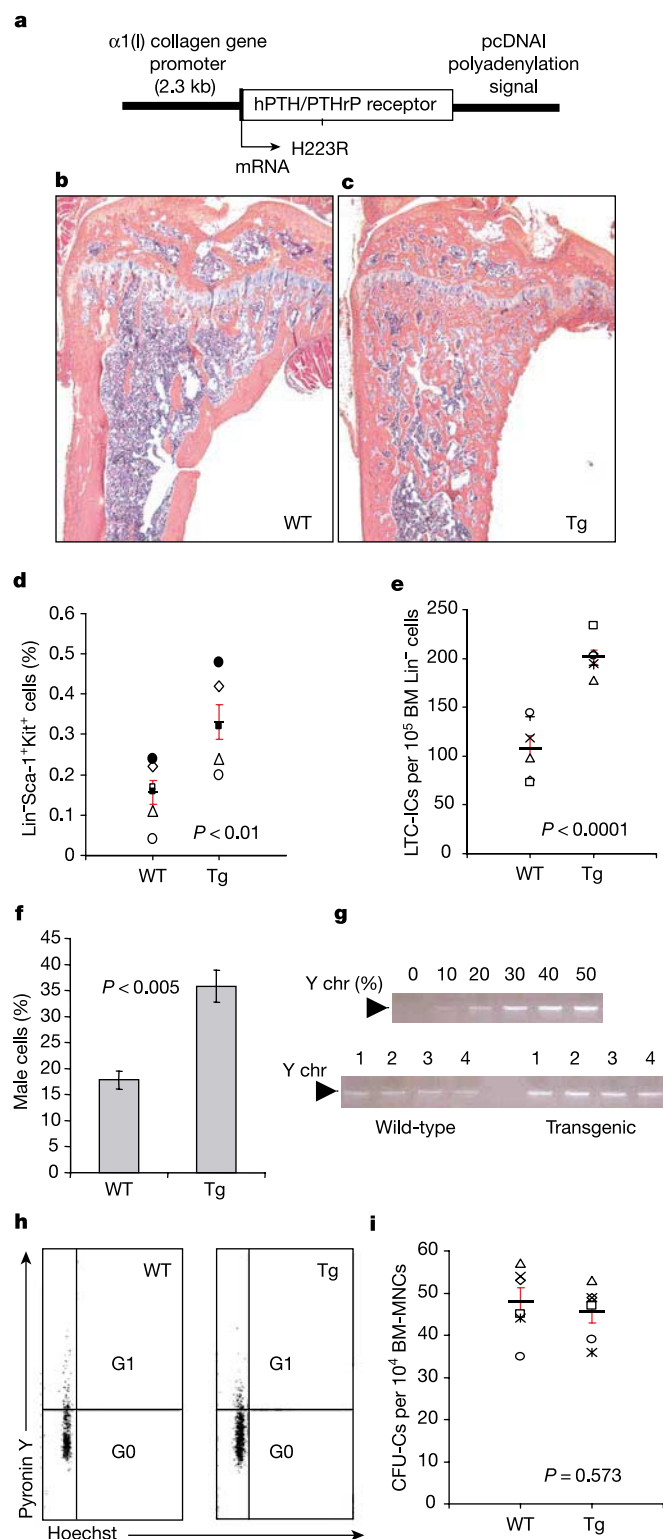


Figure 1 Col1-caPPR transgenic mice have increased numbers of trabeculae and HSCs. **a**, Representation of transgene construct. **b**, **c**, Haematoxylin and eosin staining of sections of decalcified proximal tibia from 12-week-old wild-type (WT) and transgenic (Tg) littermates. Magnifications, $\times 100$. **d**, Increased frequency of $\text{Lin}^- \text{Sca-1}^+ \text{Kit}^+$ cells in transgenic bone marrow. The same symbol designates sex-matched sibling pairs ($n = 6$ in each group; solid bar, mean). **e**, Increased LTC-ICs in transgenic bone marrow Lin^- haematopoietic cells grown on wild-type stroma. **f**, **g**, Bone marrow cells from col1-caPPR male mice had an increased ability to reconstitute female recipients by PCR for the Y chromosome ($n = 4$ in each group). **h**, No difference in HSCs residing in G0 and G1 (representative analysis from three separate experiments). **i**, Unchanged CFU-Cs.

$\alpha 1(I)$ collagen promoter active in osteoblastic cells in a transgenic mouse⁴ (Fig. 1a).

The long bones of col1-caPPR mice had increased numbers of trabeculae (Fig. 1b, c) and trabecular osteoblastic cells, as defined by expression of osteocalcin, alkaline phosphatase, collagen type I, osteopontin and matrix metalloproteinase 13 (ref. 4). Haematopoietic cells were found in small regions between trabeculae. The bony change was restricted to the metaphyseal area and there was no substantial alteration in the overall haematopoietic cell content of the bone marrow (Supplementary Fig. 1).

Assessing the composition of the bone marrow mononuclear cell fraction (BMMC), we found that the stem-cell-enriched lineage-negative (Lin^-) $\text{Sca-1}^+ \text{c-Kit}^+$ subpopulation of cells was significantly greater in the bone marrow of the transgenic animals ($P = 0.008$; Fig. 1d). There was a corresponding increase in the absolute number per hind limb: wild-type, $32,500 \pm 8,000$; transgenic, $65,700 \pm 7,500$. Testing whether functionally defined cell types were also increased, we used the quantitative limiting-dilution 'long-term culture-initiating cell' (LTC-IC) assay, which linearly correlates with *in vivo* HSC function¹². Using the Lin^- fraction of BMMC as the input population, we confirmed an increase in the frequency of primitive HSC-like cells in the transgenic animals ($P = 0.00009$; Fig. 1e). An increase in HSCs was further corroborated by competitive transplantation of irradiated recipients that showed superior engraftment of transgenic cells ($P = 0.001$; Fig. 1f, g). The magnitude of the increase was comparable in the three independent stem cell assays used. As functional assays may be influenced by the cell cycle, we analysed the proportions of $\text{Sca-1}^+ \text{Lin}^-$ cells that were in the G0 versus G1 phase, and detected no differences between the transgenic and wild-type mice ($P = 0.768$; Fig. 1h). These data confirm an increase in the stem cell population in the transgenic animals. However, when the more mature progenitor population was quantified using the CFU-C (colony forming unit-culture) assay, no difference between the transgenic and wild-type animals was observed ($P = 0.573$; Fig. 1i). Therefore, cell expansion was not global across differentiated subsets, but was notably restricted to stem cells.

To assess whether the impact on primitive cells was dependent on stroma or haematopoietic cells, we evaluated the ability of stromal cells from wild-type or transgenic animals to support wild-type BMMC in LTC-IC assays. Stromal cells expressing the transgene demonstrated augmented support of LTC-IC ($P = 0.004$; Fig. 2a). So, the increased number of primitive cells in the col1-caPPR mice was stroma-determined and was independent of the haematopoietic cell genotype.

To assess possible mechanisms, we examined levels of the cytokines interleukin 6 (IL-6), kit ligand or stem cell factor (SCF), and the chemokine stroma-derived factor 1 (SDF-1 or CXCL12), by immunohistochemistry in cells of osteoblast origin determined by osteopontin expression (Supplementary Fig. 2). Each protein was produced at increased levels in the transgenic stroma compared with wild-type controls. To address whether these diffusible cytokines could account for the effect on primitive cells, we performed LTC-ICs across a semi-permeable membrane and noted abolition of benefit from the activated PPR (Supplementary Fig. 3). Therefore, membrane- or matrix-bound cytokines were necessary for the stromal effect on HSCs.

The Notch signalling pathway involves membrane-bound ligands and regulates cell-fate specification in a wide variety of systems, including HSC self-renewal^{13–17}. By changing the balance of daughter cell self-renewal versus differentiation, Notch signalling has been shown to increase stem cell numbers without expanding mature cells in a manner reminiscent of the phenotype observed here^{13,16}. Furthermore, the Notch ligand jagged 1 (Jag1) has been shown to be expressed by marrow stromal cells^{16,18} and by murine osteoblasts¹⁹. Notch activation plus cytokine receptor signalling has a combinatorial effect on haematopoietic cells¹⁴. We therefore investigated by

immunohistochemistry Jag1 protein levels in the marrow of transgenic mice, and observed that overall levels of Jag1 were markedly increased (Fig. 2b). The cells expressing Jag1 were osteoblastic, as indicated by morphology and by osteopontin expression (Fig. 2c). To examine whether the haematopoietic stem cells responded to the increased expression of Jag1 in the transgenic animals, we assessed the level of the Notch1 intracellular domain (NICD) in the $\text{Lin}^- \text{Sca-1}^+ \text{c-Kit}^+$ HSCs from wild-type and transgenic mice. This anti-NICD antibody has previously been shown to preferentially detect the activated intracellular form of Notch1 (ref. 20). Whereas wild-type mice had minimal staining for NICD compared with isotype controls, $\text{Lin}^- \text{Sca-1}^+ \text{c-Kit}^+$ cells from transgenic mice had a notable increase in the level of NICD (Fig. 2d). Moreover, long-term co-cultures performed in the presence or absence of a γ -secretase inhibitor capable of blocking Notch cleavage²¹ showed that addition of the inhibitor reduced the supportive capacity of transgenic stroma to baseline levels (Fig. 2e). These data support a model in which PPR activation in the osteoblastic population increases cell number and the overall production of Jag1. This, in turn, may activate Notch on primitive haematopoietic cells, resulting in expansion of the stem cell compartment.

Since the col1-caPPR mice represented a genetic means of activating a receptor that could also be activated by endogenous ligand, we tested whether the effects of col1-caPPR stroma could be recapitulated through exposure of wild-type stroma to PTH. LTC-IC assays were performed using C57Bl/6 stroma expanded *in vitro* in the presence or absence of PTH, after which BMMC was introduced to the stroma in the presence or absence of PTH. When stroma was grown in medium containing PTH, it closely resembled the LTC-IC results using the col1-caPPR stroma, increasing LTC-IC ($P = 0.004$; Fig. 3a). Of note, the effect was not seen using stromal cells that were expanded in the absence of PTH, or when PTH was added at the same time as the haematopoietic cells, suggesting an effect on the composition or activity of the stroma as it matures *in vitro*. To assess whether there was an increase in the number of osteoblastic cells in the stromal cell cultures, we performed alkaline phosphatase staining. After 14 days, the cultures were confluent and heterogeneous, and there was an increase in the number of alkaline-phosphatase-positive cells in the PTH-treated cultures (Fig. 3b). These data confirm that activation of PPR induces an increase in the number of osteoblastic cells. To further assess whether the effects of PPR activation on primitive haematopoietic cells were due to

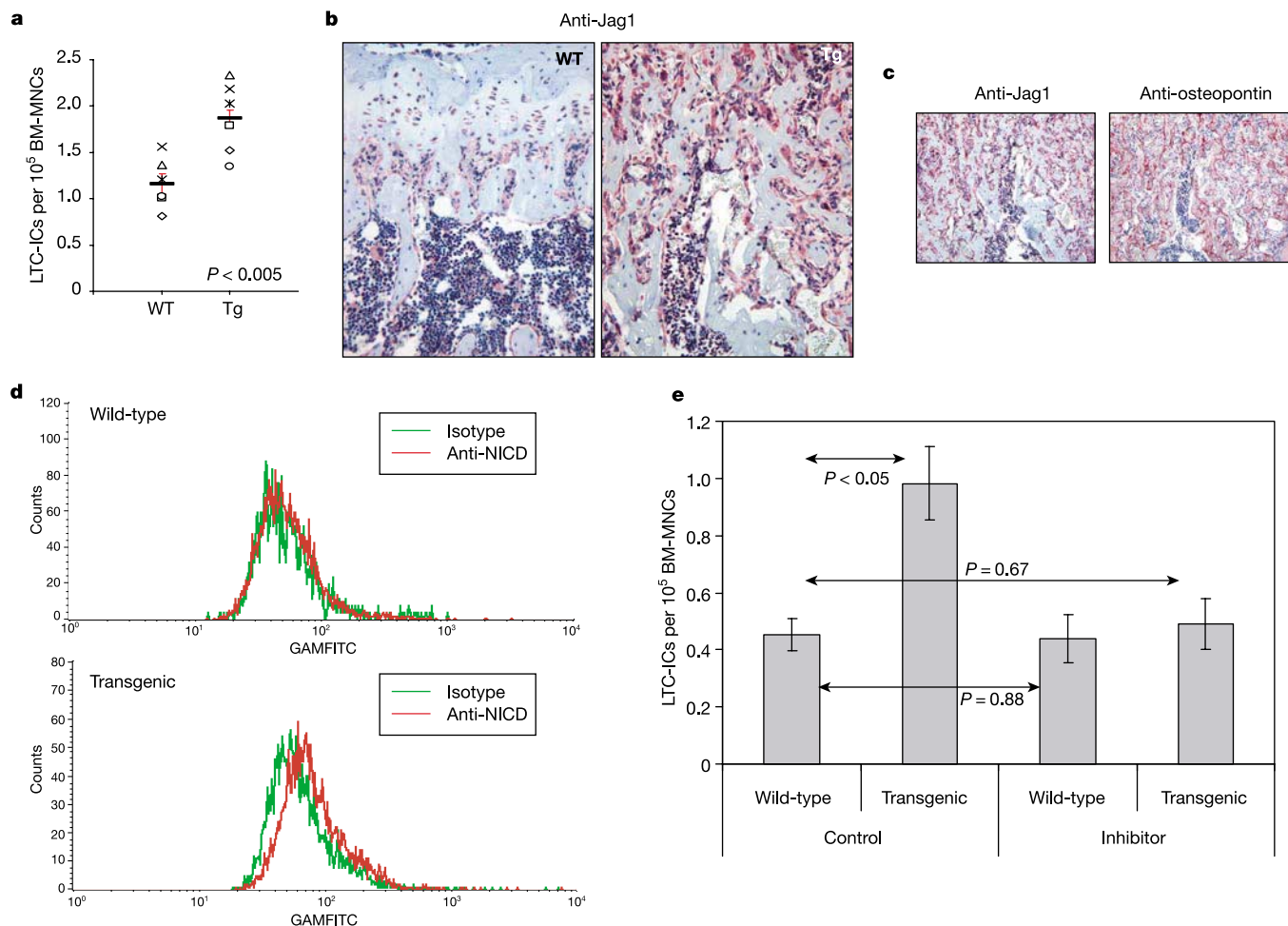


Figure 2 PPR activation of stromal cells increases HSC support and upregulates Notch signalling molecules. **a**, Stromal cells from transgenic mice ($n = 6$ versus $n = 6$ wild-type littermates) improve support of wild-type HSCs. **b**, Increased anti-Jag1 antibody staining (red) in sections of the proximal tibia from transgenic versus wild-type littermates. Magnifications, $\times 200$. **c**, Positive anti-osteopontin antibody staining in transgenic cells expressing Jag1 in sequential representative sections of the proximal

tibia from a transgenic mouse. **d**, Increased anti-NICD antibody staining of haematopoietic $\text{Lin}^- \text{Sca-1}^+ \text{c-Kit}^+$ cells from transgenic compared with wild-type littermates. Isotype controls are also shown (representative histogram from three independent experiments). **e**, HSC expansion mediated by col1-caPPR stromal cells was blocked by the addition of γ -secretase inhibitor to LTC-IC assays.

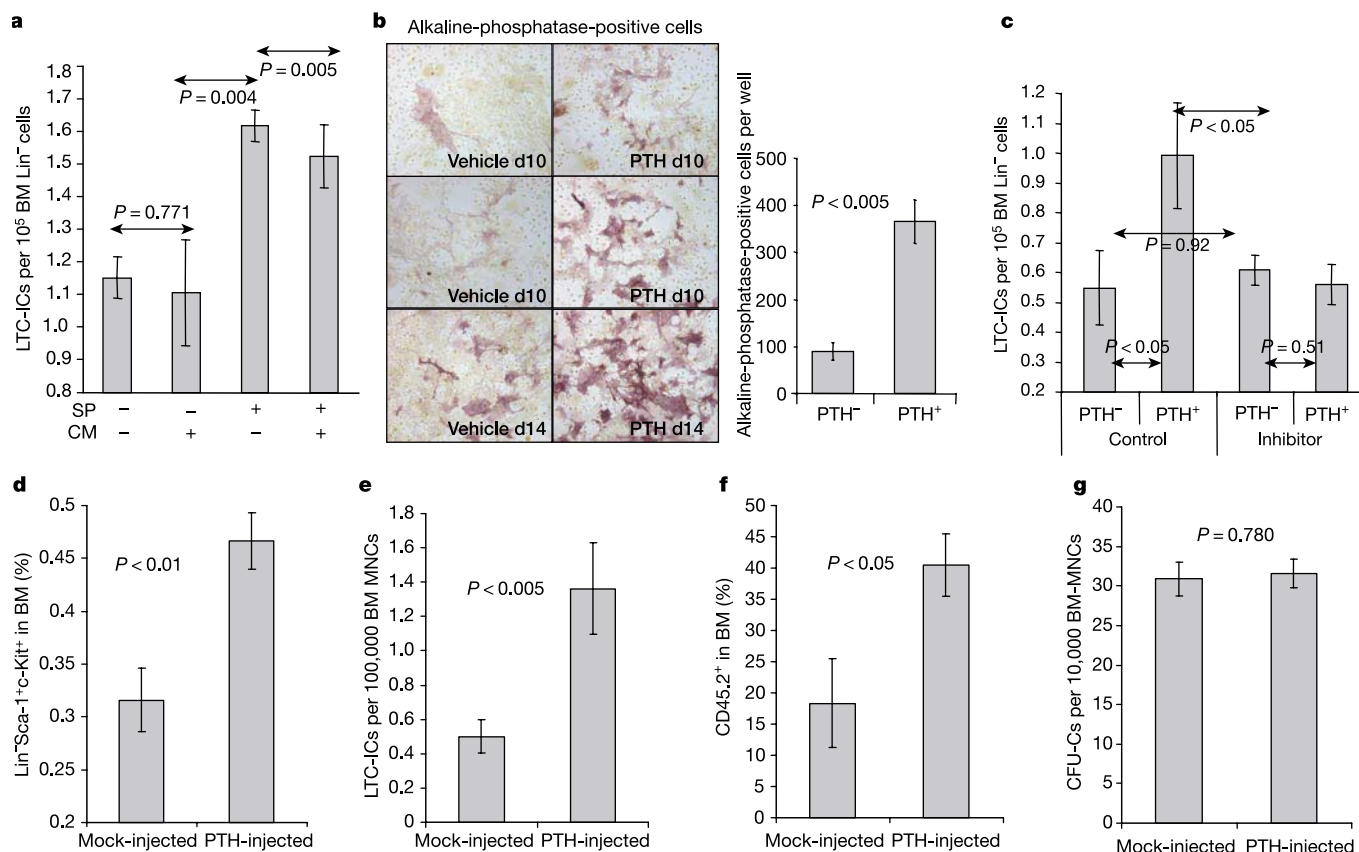


Figure 3 PTH administration expands HSCs. **a**, PTH-treated wild-type BM-SCs reproduce the col1-caPPR effect (SP, stromal preparation; CM, culture maintenance). **b**, Increased numbers of alkaline-phosphatase-positive BM-SCs (red) in PTH-treated cultures ($n = 4$ in both groups). **c**, PTH-mediated HSC expansion was blocked by the addition of γ -secretase inhibitor to LTC-IC assays. **d**, PTH-treated mice ($n = 8$) had increased Lin⁻Sca-1⁺c-Kit⁺ cells (mock-injected mice, $n = 7$) (mean absolute number

per hind limb: PTH-treated, $211,000 \pm 12,000$; mock-treated, $139,000 \pm 29,000$). **e**, PTH-treated mice had increased LTC-IC frequency. **f**, Bone marrow cells from PTH-treated mice had increased ability to reconstitute secondary recipients as shown by flow cytometry for CD45.2 ($n = 5$ in each group). **g**, Unchanged CFU-C frequency ($n = 5$ in each group).

activation of the Notch pathway, we performed LTC-IC in the presence or absence of the γ -secretase inhibitor. Addition of the inhibitor abrogated the impact of PTH on the stromal support of primitive haematopoietic cells (Fig. 3c). Therefore, Notch activation is necessary for osteoblastic-cell-induced increases in primitive haematopoietic cells. Taken together, these results support the concept that PPR activation can increase the number of osteoblastic cells, resulting in Notch1-mediated expansion of HSCs.

As the addition of exogenous PTH could influence *in vitro* stem cell support, we evaluated the pharmacological use of PTH *in vivo*. Wild-type C57Bl/6 mice were injected daily with PTH using a dosing schedule known to increase osteoblasts. After four weeks, PTH-treated mice had a significant increase in the absolute number of Lin⁻Sca-1⁺c-Kit⁺ cells ($P = 0.0049$; Fig. 3d) and LTC-IC ($P = 0.0012$; Fig. 3e) compared with vehicle-injected controls. To further clarify that functional HSCs were increased by PTH, competitive transplantation into secondary irradiated recipients was used and an increase in engraftment was documented ($P = 0.022$; Fig. 3f). These data provide evidence for an increase in the number of HSCs induced by PTH. Consistent with observations in the transgenic animals, PTH treatment did not affect the concentration of haematopoietic progenitors as assessed by CFU-C ($P = 0.780$; Fig. 3g). Therefore, pharmacological activation of PPR increased HSC, but appeared to do so without a broad haematopoietic cell expansion. These data are consistent with HSC expansion by enhanced self-renewal, a phenomenon known to result from

Notch activation^{13–17}.

To assess whether PPR stimulation could have a meaningful physiological effect, we chose a model relevant to the clinical use of stem cells in humans. PTH was administered to animals undergoing myeloablative bone marrow transplantation using limiting numbers of donor cells to mimic a setting of therapeutic need. Survival at 28 days in control mice that received mock injections after transplant was 27%. In sharp contrast, animals receiving pulse dosing of PTH had improved outcomes with 100% survival (Fig. 4a). The bone marrow histology of the two groups was also substantially different, with an increase in cellularity and a decrease in fat cells in the PTH-treated group (Fig. 4b–e). Therefore, PTH alteration of the bone marrow microenvironment can result in improved engraftment and thereby markedly affect the results of this physiological challenge. Manipulating the microenvironment is a means to influence haematopoietic outcomes that is relevant for human medicine.

The studies reported here and in anatomical detail in the accompanying paper by Zhang *et al.*²² indicate that osteoblastic cells represent a regulatory component of the bone marrow microenvironment. Zhang and colleagues show that the number of these cells is influenced by bone morphogenetic protein (BMP) signalling and directly correlates with HSC number. We show that PPR signalling affects osteoblastic cells, leading to an increase in the number of HSCs through Notch activation. The concept that a stable, custom microenvironment or niche might control stem cells

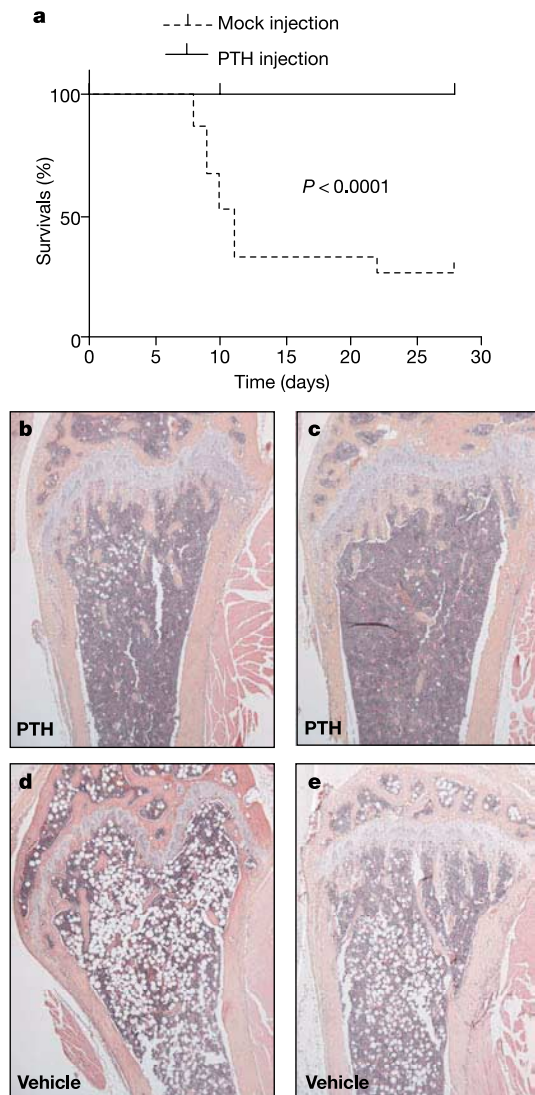


Figure 4 PTH administration confers a survival advantage after bone marrow transplantation. **a**, Kaplan–Meier survival analysis of C57Bl/6 mice transplanted with limiting numbers of BMSC cells and treated with either PTH or mock injection for four weeks ($n = 18$ in each group from three independent experiments). Histology at four weeks of long bones from either PTH-treated animals (**b**, **c**) or mock-injected controls (**d**, **e**).

was initially hypothesized by Schofield²³. In a niche, neighbouring subsets of cells and extracellular substrates house stem cells, and provide specialized functions to modulate stem cell self-renewal and progeny production *in vivo*¹. Apart from gonadal tissue of the *Drosophila*^{2,3}, physiological niches have not been well defined, largely owing to limited definition of functional microenvironment constituents. That the osteoblast may have a role in the haematopoietic system was first proposed by Taichman and Emerson^{9,24}, a hypothesis for which our data and those of Zhang *et al.*²² now provide direct *in vivo* support. By identifying cellular participants that affect stem cells *in trans*, it might be possible to define the complexity of the niche and to tease apart the systems that direct stem cell behaviour. In so doing, strategies for altering stem cells to achieve therapeutic outcomes become feasible. We show here the potential for such an approach—the definition of the osteoblast as a participant has allowed the identification of a pharmacological manipulation for expanding primitive haematopoietic cells. PTH

may be useful in stem cell expansion *ex vivo* or *in vivo*, and has potential implications for stem cell harvesting and recovery after transplantation. Understanding the interactions between bone and bone marrow may, therefore, yield practical methods for manipulating stem cells and define a model for the impact of microenvironment on cell physiology. □

Methods

Sample preparation and histological analysis

Col1-*caPPR* mice were generated and genotyped previously, as described⁴. For all studies, mice were killed by CO₂ asphyxiation. The institutional animal care committee approved all studies described.

Hind limbs from 12-week-old transgenic mice and sex-matched wild-type littermates were fixed, decalcified and processed as described⁴. For immunohistochemistry, sections were stained with anti-IL-6 goat polyclonal antibody M-19 (1:100 dilution), anti-SCF goat polyclonal antibody G-19 (1:100), anti-SDF-1 goat polyclonal antibody C-19 (1:50), anti-osteopontin goat polyclonal antibody P-18 (1:200), and anti-Jag1 rabbit polyclonal antibody H-114 (1:100) (all Santa Cruz Biotechnology, Inc.), a biotinylated rabbit anti-goat or goat anti-rabbit secondary antibody (Vector Labs), horseradish-peroxidase-conjugated streptavidin (Jackson Immuno Research), and AEC chromogen (Biocare Medical) or the Vector ABC alkaline phosphatase kit (Vector Labs). Slides were counterstained with Mayer's haematoxylin. Staining was completely absent in identical tissue sections in which the primary antibody was omitted (data not shown).

Preparation of bone marrow stromal layers

Femurs and tibias removed from euthanized mice were flushed with long-term culture medium (LTM), and bone marrow mononuclear cells (BM-MNCs) were cultured at an initial concentration of 5×10^6 cells ml⁻¹ as described²⁵.

Flow cytometric analysis

To immunophenotypically enumerate HSCs, single-cell suspensions of BMSC were stained with biotinylated lineage antibodies (anti-CD3, anti-CD4, anti-CD8, anti-TER119, anti-Gr-1, anti-Mac-1 and anti-B220), anti-Sca-1 and anti-c-Kit (Pharmingen), then labelled with a secondary fluorescent-conjugated streptavidin and analysed on a FACScalibur cytometer (Becton Dickinson and Co.) using Cell Quest software. To assess the cell cycle in the primitive population, BMSC cells were stained with lineage antibodies, anti-Sca-1, Pyronin Y (RNA dye) and Hoechst 33342 (DNA dye) as described²⁵. For intracellular NICD staining, Lin⁻Sca-1⁺c-Kit⁺ cells were permeabilized using the Fix and Perm cell-permeabilization kit (Caltag) and incubated with 1 µg of anti-NICD antibody and secondary goat anti-mouse antibody.

Colony-forming unit assay

Isolated BM-MNCs were assessed for CFU-C frequency as described previously²⁵. In this study, we used the methylcellulose-containing medium M3434 (StemCell Technologies).

Long-term culture-initiating cell assay

LTC-IC frequency in Lin⁻ BMSC preparations was quantified as described previously¹³.

In vitro treatment with PTH

LTC-IC assays were performed using wild-type stroma and haematopoietic cells. Rat PTH (1–34) (Bachem) or vehicle was added to each media change, either during stroma establishment and/or during culture maintenance (PTH final concentration, 10^{-7} M). For alkaline phosphatase staining, primary bone marrow stromal cells (BM-SCs) were cultured in 24-well plates. After 10 or 14 days, the adherent cells were fixed in 10% neutral formalin buffer, and alkaline phosphatase activity was determined histochemically by incubation with 0.1 mg ml⁻¹ naphthol AS-MX phosphate (Sigma), 0.5% *N,N*-dimethylformamide, 0.6 mg ml⁻¹ red violet LB salt (Sigma) in 0.1 M Tris-HCl (pH 8.5). To inhibit γ -secretase activity, 30 µM of γ -secretase inhibitor II (Calbiochem) in dimethylsulphoxide was added to the LTM and LTC-IC assays were performed. For non-contact LTC-ICs, BM-SC layers were plated into 96-well plates as described. BMSC cells were seeded into culture inserts with a membrane of 0.2 µm pore size (Nunc) placed in the wells.

In vivo PTH administration

Rat PTH (1–34) (Bachem) (80 µg per kg body weight) or vehicle was injected intraperitoneally five times a week for four weeks in 6–8-week-old wild-type C57Bl/6 male mice.

Bone marrow transplantation

For *in vivo* quantification of HSCs from wild-type and transgenic mice, 1×10^7 BM-MNCs from female FVB mice (Jackson Laboratories) were mixed with 5×10^6 cells from male wild-type or transgenic mice. Cells were injected into recipient female FVB mice lethally irradiated 24 h previously with 8.5 Gy of radiation (¹³⁷Cs source). After 8 weeks, the bone marrow was removed and flushed with fully supplemented Iscove's medium. The relative contribution of engraftment from the different cell sources was assessed by semiquantitative polymerase chain reaction (PCR) for Y chromosome using primers and conditions as described previously²⁶. Band density on the agarose gel was quantified using a ChemImager 5500 (Alpha Innotech).

For *in vivo* quantification of HSCs following PTH administration, 4×10^5 BMSC cells from CD45.1 + B6.SJL mice (Jackson Laboratories) were mixed with 2×10^5 cells from

vehicle-injected or PTH-injected CD45.2⁺ C57Bl/6 mice. Cells were injected in recipient B6.SJL mice lethally irradiated 24 h previously with 10 Gy of radiation. The relative contribution of engraftment from the different cell sources was assessed by flow cytometry of BM-MNCs using anti-CD45.1 and anti-CD45.2 antibodies (Pharmingen).

To assess post-transplantation PTH effects, lethally irradiated recipient C57Bl/6 mice were injected with 2×10^5 BM-MNCs from a donor B6.SJL mouse. Twenty-four hours later, mice were injected with PTH or vehicle for four weeks.

Statistical analysis

Results are expressed as mean $\pm$ s.e.m. Data were analysed using the unpaired two-tailed Student's *t*-test as appropriate for the data set. $P < 0.05$ was considered significant.

Received 30 May; accepted 26 August 2003; doi:10.1038/nature02040.

1. Spradling, A., Drummond-Barbosa, D. & Kai, T. Stem cells find their niche. *Nature* **414**, 98–104 (2001).
2. Kiger, A. A., White-Cooper, H. & Fuller, M. T. Somatic support cells restrict germline stem cell self-renewal and promote differentiation. *Nature* **407**, 750–754 (2000).
3. Tran, J., Brenner, T. J. & DiNardo, S. Somatic control over the germline stem cell lineage during *Drosophila* spermatogenesis. *Nature* **407**, 754–757 (2000).
4. Calvi, L. M. *et al.* Activated parathyroid hormone/parathyroid hormone-related protein receptor in osteoblastic cells differentially affects cortical and trabecular bone. *J. Clin. Invest.* **107**, 277–286 (2001).
5. Lord, B. I., Testa, N. G. & Hendry, J. H. The relative spatial distributions of CFUs and CFUc in the normal mouse femur. *Blood* **46**, 65–72 (1975).
6. Gong, J. K. Endosteal marrow: a rich source of hematopoietic stem cells. *Science* **199**, 1443–1445 (1978).
7. Cui, Y. F., Lord, B. I., Woolford, L. B. & Testa, N. G. The relative spatial distribution of *in vitro*-CFUs in the bone marrow, responding to specific growth factors. *Cell Prolif.* **29**, 243–257 (1996).
8. Lord, B. I. The architecture of bone marrow cell populations. *Int. J. Cell Cloning* **8**, 317–331 (1990).
9. Taichman, R. S. & Emerson, S. G. Human osteoblasts support hematopoiesis through the production of granulocyte colony-stimulating factor. *J. Exp. Med.* **179**, 1677–1682 (1994).
10. Taichman, R. S., Reilly, M. J. & Emerson, S. G. Human osteoblasts support human hematopoietic progenitor cells *in vitro* bone marrow cultures. *Blood* **87**, 518–524 (1996).
11. Taichman, R., Reilly, M., Verma, R., Ehrenman, K. & Emerson, S. Hepatocyte growth factor is secreted by osteoblasts and cooperatively permits the survival of haematopoietic progenitors. *Br. J. Haematol.* **112**, 438–448 (2001).
12. Ploemacher, R. E., van der Sluijs, J. P., van Beurden, C. A., Baert, M. R. & Chan, P. L. Use of limiting-dilution type long-term marrow cultures in frequency analysis of marrow-repopulating and spleen colony-forming hematopoietic stem cells in the mouse. *Blood* **78**, 2527–2533 (1991).
13. Stier, S., Cheng, T., Dombkowski, D., Carlesso, N. & Scadden, D. T. Notch1 activation increases hematopoietic stem cell self-renewal *in vivo* and favors lymphoid over myeloid lineage outcome. *Blood* **99**, 2369–2378 (2002).
14. Varnum-Finney, B., Brashem-Stein, C. & Bernstein, I. D. Combined effects of Notch signalling and cytokines induce a multiple log increase in precursors with lymphoid and myeloid reconstituting ability. *Blood* **101**, 1784–1789 (2003).
15. Varnum-Finney, B. *et al.* Pluripotent, cytokine-dependent, hematopoietic stem cells are immortalized by constitutive Notch1 signalling. *Nature Med.* **6**, 1278–1281 (2000).
16. Karanu, F. N. *et al.* The notch ligand jagged-1 represents a novel growth factor of human hematopoietic stem cells. *J. Exp. Med.* **192**, 1365–1372 (2000).
17. Karanu, F. N. *et al.* Human homologues of Delta-1 and Delta-4 function as mitogenic regulators of primitive human hematopoietic cells. *Blood* **97**, 1960–1967 (2001).
18. Li, L. *et al.* The human homolog of rat Jagged1 expressed by marrow stroma inhibits differentiation of 32D cells through interaction with Notch1. *Immunity* **8**, 43–55 (1998).
19. Pereira, R. M., Delany, A. M., Durant, D. & Canalis, E. Cortisol regulates the expression of Notch in osteoblasts. *J. Cell. Biochem.* **85**, 252–258 (2002).
20. Huppert, S. S. *et al.* Embryonic lethality in mice homozygous for a processing-deficient allele of Notch1. *Nature* **405**, 966–970 (2000).
21. Wolfe, M. S. *et al.* Peptidomimetic probes and molecular modeling suggest that Alzheimer's γ -secretase is an intramembrane-cleaving aspartyl protease. *Biochemistry* **38**, 4720–4727 (1999).
22. Zhang, J. *et al.* Identification of the haematopoietic stem cell niche and control of the niche size. *Nature* **425**, 836–841 (2003).
23. Schofield, R. The relationship between the spleen colony-forming cell and the haematopoietic stem cell. *Blood Cells* **4**, 7–25 (1978).
24. Taichman, R. S., Reilly, M. J. & Emerson, S. G. The hematopoietic microenvironment: osteoblasts and the hematopoietic microenvironment. *Hematology* **4**, 421–426 (2000).
25. Cheng, T. *et al.* Hematopoietic stem cell quiescence maintained by p21cip1/waf1. *Science* **287**, 1804–1808 (2000).
26. Giri, N., Kaushiva, A., Wu, T., Sellers, S. E. & Tisdale, J. F. The effects of SCF/G-CSF prestimulation on radiation sensitivity and engraftment in nonmyeloablative murine hosts. *Exp. Hematol.* **29**, 779–785 (2001).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements The authors thank J. Harvey for assistance with histological analyses. This work was supported by the National Institutes of Health (L.M.C., H.M.K., E.S. and D.T.S.), the American Society of Hematology Fellow Scholar Award (G.B.A.) and the Doris Duke Foundational Burroughs Wellcome Fund (D.T.S.).

Competing interests statement The authors declare competing financial interests: details accompany the paper on www.nature.com/nature.

Correspondence and requests for materials should be addressed to D.T.S. (scadden.david@mgm.harvard.edu).

Widespread requirement for Hedgehog ligand stimulation in growth of digestive tract tumours

David M. Berman^{1,2,3*}, Sunil S. Karhadkar^{1,2*}, Anirban Maitra^{2*}, Rocio Montes de Oca¹, Meg R. Gerstenblith¹, Kimberly Briggs³, Antony R. Parker², Yutaka Shimada¹, James R. Eshleman², D. Neil Watkins³ & Philip A. Beachy¹

¹Department of Molecular Biology and Genetics and Howard Hughes Medical Institute, ²Department of Pathology, and ³Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA

⁴Department of Surgery, Kyoto University, Kyoto 606-8507, Japan

* These authors contributed equally to this work

Activation of the Hedgehog (Hh) signalling pathway by sporadic mutations or in familial conditions such as Gorlin's syndrome is associated with tumorigenesis in skin, the cerebellum and skeletal muscle^{1,2}. Here we show that a wide range of digestive tract tumours, including most of those originating in the oesophagus, stomach, biliary tract and pancreas, but not in the colon, display increased Hh pathway activity, which is suppressible by cyclopamine, a Hh pathway antagonist. Cyclopamine also suppresses cell growth *in vitro* and causes durable regression of xenograft tumours *in vivo*. Unlike in Gorlin's syndrome tumours, pathway activity and cell growth in these digestive tract tumours are driven by endogenous expression of Hh ligands, as indicated by the presence of *Sonic hedgehog* and *Indian hedgehog* transcripts, by the pathway- and growth-inhibitory activity of a Hh-neutralizing antibody, and by the dramatic growth-stimulatory activity of exogenously added Hh ligand. Our results identify a group of common lethal malignancies in which Hh pathway activity, essential for tumour growth, is activated not by mutation but by ligand expression.

The Hh signalling pathway specifies patterns of cell growth and differentiation in a wide variety of embryonic tissues. Mutational activation of the Hh pathway, whether sporadic or in Gorlin's syndrome, is associated with tumorigenesis in a small subset of these tissues, predominantly skin, the cerebellum and skeletal muscle^{1,2}. Mutations that activate the Hh pathway include those that impair the ability of the transporter-like Hh receptor³ Patched (PTCH), the target of Gorlin's syndrome mutations) to restrain Smoothened (SMO)-mediated activation of transcriptional targets through the Gli family of latent transcription factors^{1,2,4,5}. Paradoxically, Hh pathway activity is associated with increased expression of *PTCH*, which is a transcriptional target of the pathway but is unable to restrain SMO when bound by Hh protein. Pathway activation, whether triggered by Hh binding or by *PTCH* mutation, requires SMO, a seven-transmembrane protein that binds to and is inactivated by the pathway antagonist cyclopamine⁶.

The recent finding that Hh pathway activity is important for growth of a significant proportion of small-cell lung cancers⁷, a tumour type not associated with Gorlin's syndrome, suggested that other, non-Gorlin's tumours might require Hh pathway activity for growth. We investigate here the role of pathway activity in tumours derived from the gut, a tissue with prominent and diverse roles for Hh signalling in developmental patterning, and in mature tissue homeostasis. A role in homeostasis is suggested by the expression of Hh ligands and target genes in postnatal gut epithelium and mesenchyme^{8–11} (Fig. 1a).

We began our examination of gut-derived tumours by testing for expression of *Sonic hedgehog* (SHH) and *Indian hedgehog* (IHH), which encode members of the Hh ligand family that are expressed in

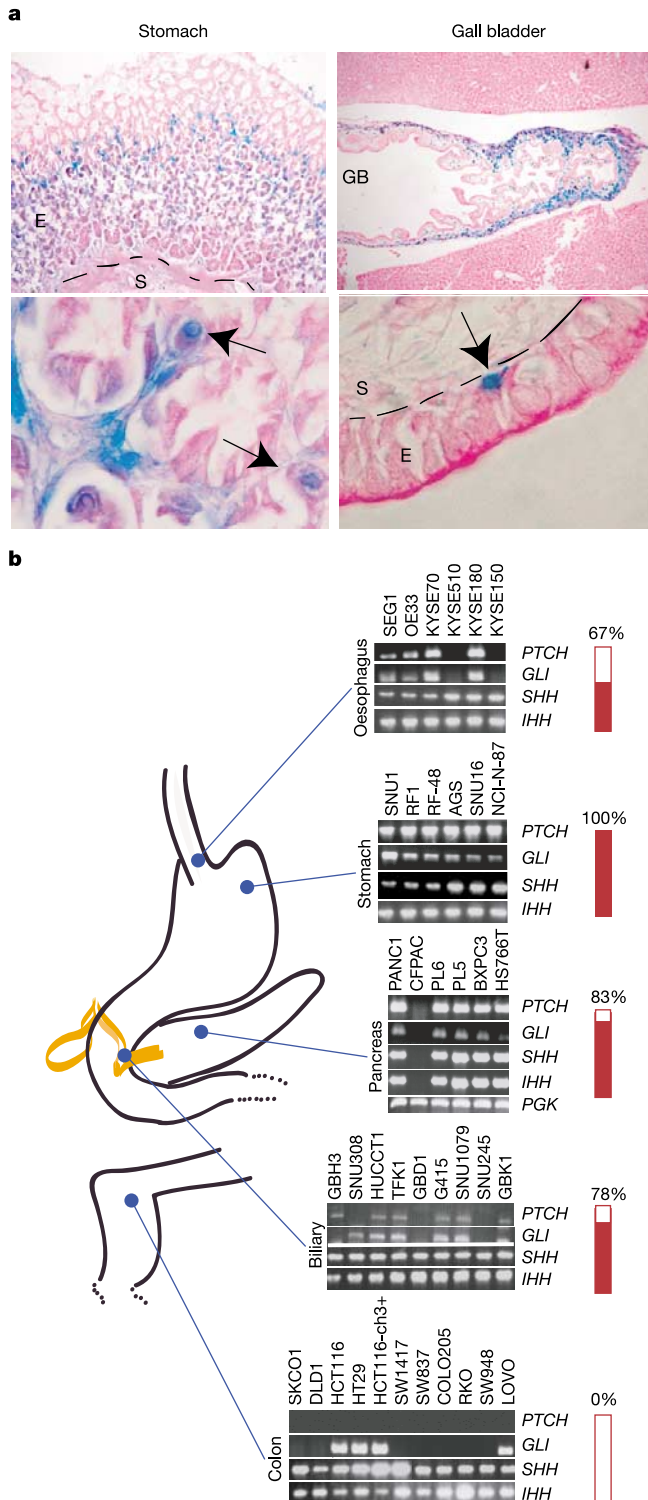


Figure 1 Hh pathway activity in normal and neoplastic gut cells and tissues. **a**, Intra-epithelial Hh pathway activation in mature murine gut tissues. Staining for β -galactosidase activity (blue) from the *Ptch-lacZ* gene reveals expression of the Hh target *Ptch* in epithelium (E) of the stomach (arrows indicate apparent parietal cells) and gall bladder (GB; arrow indicates basally oriented cell). Staining within stromal cells (S) was also seen. **b**, Widespread expression of transcripts encoding Hh pathway components in digestive tract tumour cell lines. RT-PCR products showing expression of genes encoding Hh pathway ligands (*SHH* and *IHH*) and target genes (*PTCH* and *GLI*) in cell lines taken from sites shown in the diagram on the left. Red bars indicate the percentage of lines expressing detectable *PTCH* mRNA at each site.

early endoderm and throughout gut development^{10,12}. We detected *SHH* and *IHH* messenger RNA in 37 of 38 (97%) cell lines from oesophageal, stomach, biliary tract, pancreatic and colon carcinomas (Fig. 1b). The Hh target genes *PTCH* and *GLI* were used as indicators of Hh pathway activity and were co-expressed in most cell lines from oesophageal (4/6), stomach (6/6), pancreatic (5/6) and biliary tract (5/9) tumours. By contrast, *PTCH* was not expressed in colon tumour cell lines (0/11), although a few of these cell lines expressed *GLI* in the absence of *PTCH*. The Hh pathway, however, is not active in these colon-tumour-derived cell lines, as indicated by our reporter assays (see below).

Expression of *PTCH* and *GLI* within cells from several types of digestive tract tumours that also express Hh ligands suggests the autonomous operation of an active signalling process. Autonomous pathway activity was confirmed by the high level of luciferase activity produced by an exogenously introduced Hh-inducible GLI-luciferase reporter¹³ in all cell lines producing detectable

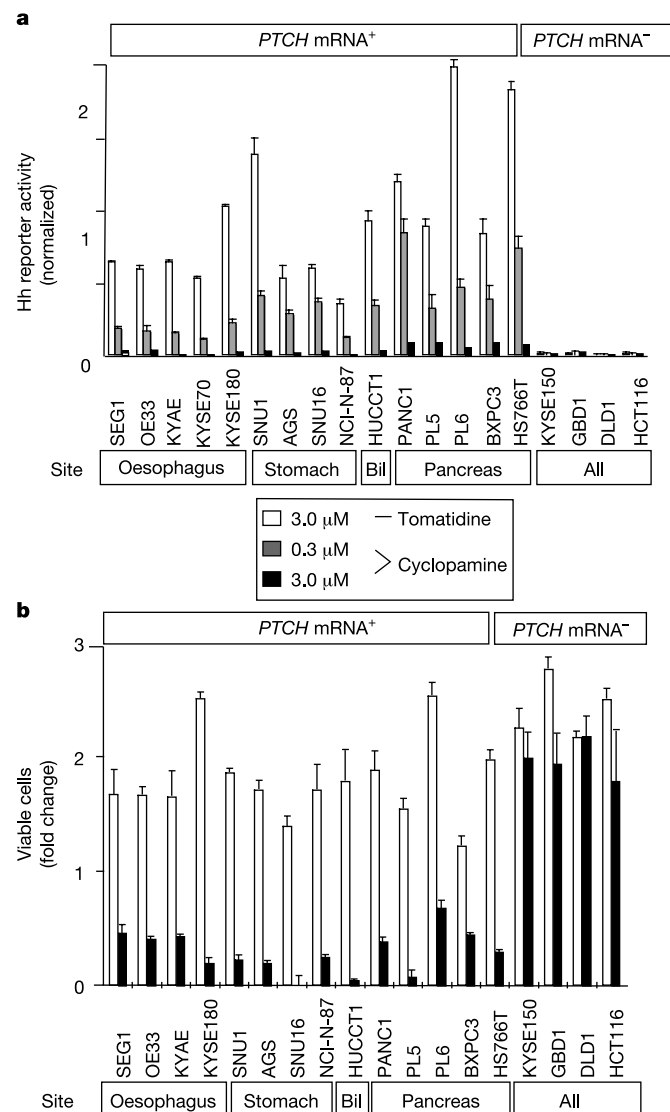


Figure 2 Cyclopamine suppression of Hh pathway activity and growth in digestive tract tumour cell lines correlates with expression of *PTCH* mRNA. **a**, Normalized activity of transiently transfected Hh-responsive luciferase reporter and dose-dependent suppression by the Hh pathway antagonist cyclopamine. **b**, Change in tumour cell viability measured by MTS (soluble tetrazolium salt) assay after culture in 3.0 μ M cyclopamine or tomatidine (control). Bil, biliary tract. Note that the viability of *PTCH* mRNA-negative cells is not affected by cyclopamine.

PTCH mRNA (Fig. 2a). Furthermore, Hh pathway activity in these cell lines was inhibited in a dose-dependent manner by the Hh-pathway-specific antagonist cyclopamine, but not by tomatidine, an inactive but structurally related compound¹⁴ (Fig. 2a). These results suggest that high levels of Hh pathway activity are a common feature of digestive tract tumours and prompted us to further investigate the role of the pathway in tumour growth. We found that cyclopamine treatment reduced the growth of tumour cell lines from the oesophagus, stomach, biliary tract and pancreas by 75–95% compared with tomatidine controls (Fig. 2b). Strikingly, significant growth inhibition was observed only in tumour lines expressing *PTCH* mRNA, confirming that the effects of cyclopamine treatment are pathway specific rather than generally cytotoxic.

Because the properties of cell lines adapted to long-term growth *in vitro* can differ from those of tumours growing *in vivo*, we also examined pathway activation in freshly resected stomach and pancreatic tumours by measuring endogenous *PTCH* mRNA levels. For each specimen, RNA for quantitative polymerase chain reaction with reverse transcription (RT-PCR) analysis was isolated from 10 consecutive 10 μ m cryosections after histological analysis of the immediately flanking sections to determine tumour content. We found that *PTCH* mRNA levels were 23–271 times (mean = 129; $n = 9$) higher in stomach tumours and 69–5,044 times (mean = 448; $n = 15$) higher in pancreatic tumours than in adjacent normal tissue (Fig. 3a).

To examine the role of Hh pathway activity in growth, we analysed pancreatic carcinomas that had been passaged once as xenografts in nude mice then cultured and immediately assayed *in vitro*. Of six such xenografts, four expressed *PTCH* mRNA (data

not shown); two of these were a matched pair of primary and metastatic tumours from a single patient. All four of these *PTCH*-expressing primary xenografts expressed the GLI-luciferase reporter in a cyclopamine-sensitive manner (Fig. 3b). Cyclopamine treatment of these *PTCH* mRNA-expressing xenografts also resulted in decreased viable cell mass (Fig. 3c), demonstrating more extreme cell-killing effects of Hh pathway blockade than those observed in established tumour cell lines (Fig. 2b). By contrast, single-passage xenografts lacking *PTCH* mRNA grew equally well in control and cyclopamine-containing media (Fig. 3c), again confirming that cyclopamine effects are pathway specific rather than generally cytotoxic.

To examine the effects of cyclopamine treatment *in vivo*, subcutaneous xenografts from HUCCT1 cells, a metastatic cholangiocarcinoma cell line, were established in athymic mice. After the tumours had grown to an average size of 180 mm³, mice bearing these tumours were injected daily with cyclopamine or vehicle, and control tumours continued to grow until animals were euthanized at 22 days with an average tumour volume exceeding 800 mm³ (Fig. 3d, e). Tumours in cyclopamine-treated animals, by contrast, regressed completely by 12 days. Treatment continued for a further 10 days, followed by an observation period of 98 days. Remarkably, all treated tumours were grossly and histologically undetectable at the end of the 3 month observation period, indicating a complete and durable tumoricidal effect of blocking the Hh pathway with cyclopamine (Fig. 3d, e). As previously reported for shorter treatment courses^{7,15}, all mice survived cyclopamine treatment and the observation period with no obvious adverse effects.

These findings establish that the Hh pathway is widely activated

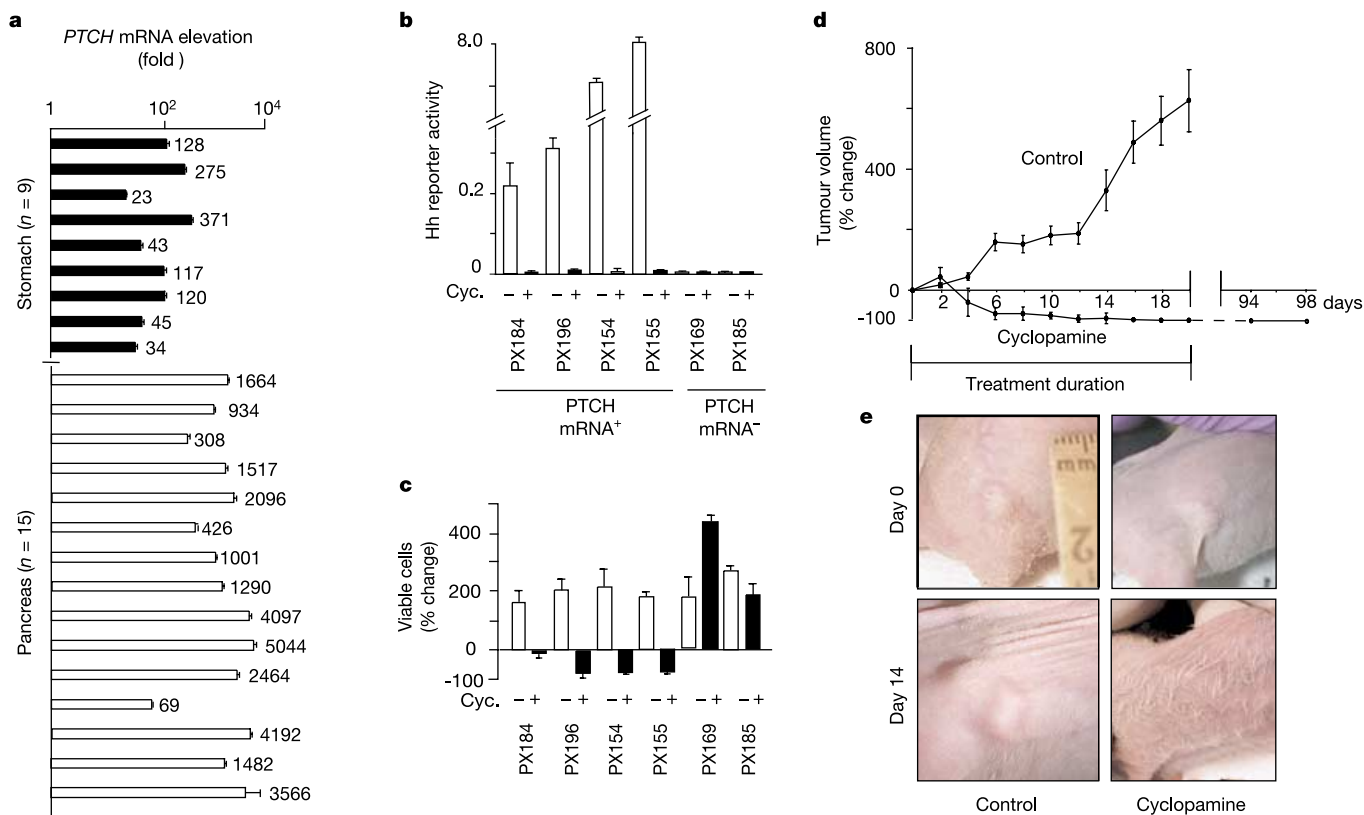


Figure 3 Hh pathway activity and requirement for growth of tumour cells *in vivo*. **a**, Elevated *PTCH* mRNA in surgically resected pancreatic and gastric carcinomas was detected by quantitative RT-PCR and normalized to adjacent normal stomach ($n = 10$) and pancreas ($n = 1$) levels. **b**, Normalized Hh-responsive reporter activity and suppression by 3.0 μ M cyclopamine in first-passage pancreas carcinoma xenografts. **c**, Corresponding reduction in viable tumour cells when cultured with 3.0 μ M

cyclopamine. Note that reduced viability is observed exclusively in lines with elevated Hh pathway activity. **d**, Change in HUCCT1 human cholangiocarcinoma xenograft volume in mice treated for 22 days with vehicle (control; $n = 9$) or cyclopamine ($n = 9$). A durable response was observed for the full 98-day observation period after cessation of therapy. **e**, Representative photographs of control and cyclopamine-treated mice. Note the full regression of the tumour on the lower right.

in gut-derived tumours, and further demonstrate a role for pathway activity in tumour cell growth *in vitro* and *in vivo*. Yet Gorlin's syndrome is not associated with a higher incidence of gut-derived tumours, and *PTCH* mutations in these tumours have not been reported, which suggests a distinct mechanism for Hh pathway activation that does not involve mutation of pathway components. Given that *SHH* and *IHH* mRNA are expressed in nearly all gut-derived tumours examined, we investigated the role of Hh ligand binding in pathway activity. We measured Hh-inducible reporter activity in cells from tumours of the oesophagus, stomach, pancreas

and biliary tract treated with 5E1 monoclonal antibody¹⁶, which binds to Shh and Ihh ligands¹⁷ and blocks signalling by disrupting ligand binding to Ptch¹⁸. Autonomous activation of the transfected reporter was not affected by the control antibody, but was markedly reduced by incubation with 5E1 at 0.1 or 10 $\mu\text{g ml}^{-1}$ (Fig. 4a and Supplementary Fig. 1a). By contrast, reporter activity was augmented by around eightfold by the addition of purified Shh ligand to a concentration of 25 nM. Addition of 5E1 in combination with Shh ligand reduced reporter activity to a level intermediate between those produced with either reagent alone (Fig. 4a and Supplemen-

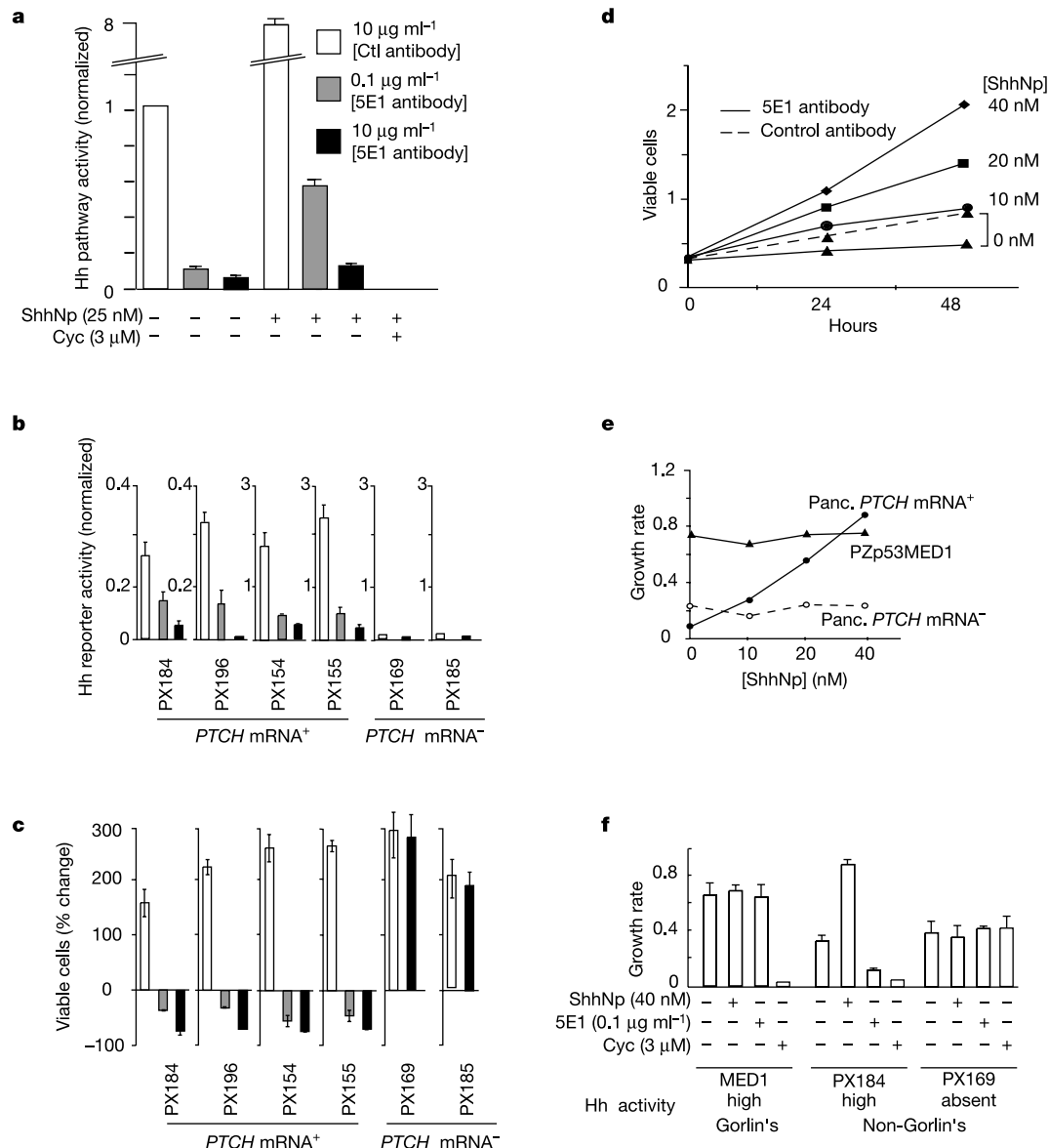


Figure 4 Ligand dependence of Hh pathway activity and growth in digestive tract tumours. **a**, Mutually antagonistic effects of the Hh ligand and blocking antibody on activity of the Hh reporter. The Hh-neutralizing 5E1 monoclonal antibody suppresses and Shh ligand increases reporter activity in HUCCT1 cells. Combined addition of antibody and ligand produces intermediate effects, depending on their relative concentrations. **b**, Hh reporter activity in first-passage pancreatic carcinoma xenografts and dose-dependent suppression with 5E1 monoclonal antibody. **c**, MTS assay showing reduced viability corresponding to Hh pathway suppression by 5E1 monoclonal antibody. **d**, MTS assay showing growth (in arbitrary units) of PX184 first-passage *PTCH*-mRNA-expressing pancreas xenograft cells cultured with the control antibody (dashed line) or with 5E1 monoclonal antibody at a level just sufficient to suppress growth (0.1 $\mu\text{g ml}^{-1}$; solid lines)

and with the indicated concentrations of added Shh ligand. **e**, **f**, Modulation of cell growth rate (in arbitrary units) by 5E1 antibody, Shh ligand and cyclopamine in single-passage pancreatic xenografts and in medulloblastoma cells (PZp53^{MED1}). **e**, Shh ligand concentrations control growth rates of *PTCH* mRNA-positive pancreatic xenograft lines (average; $n = 4$) but not of *PTCH* mRNA-negative lines ($n = 2$) or of PZp53^{MED1} cells. **f**, Opposite responses to ligand and antibody of PX-184 cells, which express *PTCH* mRNA, and the lack of response of PX-169 and PZp53^{MED1} cells, which respectively lack detectable Hh pathway activation or display constitutive pathway activation owing to lack of functional *PTCH*¹⁵. Cyclopamine, by contrast, blocks growth in cells with pathway activation because of either *PTCH* mutation or ligand stimulation.

tary Fig. 1a), which indicates a mutual antagonism between 5E1 antibody and Hh ligand in activating the pathway.

Reporter activity in cells from single-passage pancreatic cancer xenografts was also antagonized by 5E1 (Fig. 4b). In addition, treatment with the 5E1 antibody markedly reduced viable cell mass (Fig. 4c). This cell-killing effect and the reporter effect were observed exclusively in cells from tumours that expressed endogenous *PTCH* mRNA. We further investigated the relationship between ligand concentration and growth by adding 5E1 antibody to cells from a single-passage pancreatic tumour xenograft at a level that was just sufficient to block growth. We then added Shh protein and found that growth correlated positively with increasing concentrations (Fig. 4d). Rates of growth from this experiment plotted as a function of Shh concentration (Fig. 4e) indicate that ligand-induced pathway activation is rate limiting and that unperturbed growth of these cells is sub-maximal. Assays of tumour cell lines derived from the oesophagus, stomach, pancreas and biliary tract yielded similar results (Supplementary Fig. 1a–c), confirming a widespread requirement for Hh ligand in the growth of these tumours.

The Hh ligand and 5E1 antibody are mutually antagonistic in their effects on reporter activity and produce opposite effects on the growth of cells from these gut-derived tumours (Fig. 4a–e and Supplementary Fig. 1a–c). Thus, pathway activation and cell growth must be dependent on Hh ligand. By contrast, addition of neither Hh ligand nor 5E1 blocking antibody significantly affected the growth of cells from a single-passage pancreatic tumour xenograft that did not express *PTCH* mRNA (Fig. 4f), which demonstrates the specificity of antibody and ligand effects. We also observed no significant ligand- or antibody-induced change in growth of medulloblastoma cells derived from a mouse model of Gorlin's syndrome (Fig. 4f) in which the Hh pathway is activated through loss of *Ptch* function^{15,19}. In contrast to antibody-resistant xenograft cells, medulloblastoma-derived cells require pathway activity for growth and can be killed by cyclopamine treatment¹⁵ (Fig. 4f).

Ligand-independent mutational activation of the Hh pathway has been linked to the formation of tumours, such as medulloblastoma, associated with Gorlin's syndrome. Despite a widespread activation of and dependence on the Hh pathway for medulloblastoma growth¹⁵, only a fraction of sporadic tumours can be assigned to pathway-activating mutations, suggesting that other mechanisms of pathway activation may be at play. Here we establish such a mechanism by showing that pathway activation and growth of cells from a group of commonly lethal gut-derived malignancies is ligand dependent. Small-cell lung cancer, also arising from endodermally derived epithelium and associated with Hh ligand expression, has recently been linked to transient reactivation of the Hh pathway within the airway epithelium, where it regulates progenitor cell fates during injury repair⁷. A similar role for Hh signalling in renewal of mature digestive tract epithelium is suggested by expression of the Hh pathway targets *Ptch* and *Gli3*^{9,10} (Fig. 1a) and by the requirement for Hh signalling for proliferation of gut progenitor cells¹⁰. It is not known whether renewal of injured gut epithelium is associated with transient Hh pathway reactivation. However, increased rates of oesophageal, gastric and pancreatic carcinomas occur in association with acid injury in Barrett's oesophagus, in *Helicobacter pylori* infection, and with exposure to alcohol, cigarette smoke and certain dietary components^{20–22}. Exposure to such factors probably causes injury to the gut epithelium, eliciting a chronic state of injury repair and a consequent increase in proliferative stem or progenitor cells that may arise through ligand-dependent reactivation of the Hh pathway. Many of these agents are also mutagenic, thus potentially enhancing tumour formation by subjecting an enlarged pool of stem or stem-like target cells to potentially oncogenic mutations. Our results identify a group of common and frequently lethal gut-derived tumours, readily diagnosed by their expression of endogenous pathway targets such as *PTCH*, which may respond to antagonist- or

antibody-mediated pathway blockade, even at advanced stages of metastatic disease. □

Methods

Detection of β -galactosidase expression

Ptch-lacZ mice were killed at 4–6 weeks of age. Staining was performed as described previously⁷.

Tumour cells and tissues

Origins and sources of our cells and tissues are described in the Supplementary Table. First-passage pancreatic cancer xenografts were derived from freshly harvested pancreaticoduodenectomy specimens as described previously¹⁵. In previous experiments, approximately 65% of specimens yielded xenografts (data not shown), so we inferred that these xenografts represent pancreatic tumours in the general population. The diagnosis of frozen samples from gastric and pancreatic adenocarcinoma resections and adjacent normal stomach and pancreas was microscopically confirmed by two pathologists (D.M.B. and A.M.), and RNA was prepared as described previously¹⁵.

RT-PCR

Templates were prepared and amplified as described elsewhere¹⁵. For all primer pairs, specificity was confirmed by sequencing of PCR products. For quantitative RT-PCR, 10% of the first-strand reaction was amplified using IQ-SYBR Green Supermix, an i-cyclerIQ real-time detection system (Bio-Rad) and specific oligonucleotide primers for *PTCH* or *PGK* (phosphoglycerate kinase). Amplification was performed at 95 °C for 5 min followed by 40 cycles of 10, 15 and 30 s at 95 °C, 55 °C and 75 °C, respectively. Bio-Rad software was used to calculate threshold cycle (C_T) values for *PTCH* and for the housekeeping gene *PGK*. For each sample, *PTCH* expression was derived from the ratio of *PTCH* to *PGK* levels using the formula $2^{-\Delta C_T}$, where $\Delta C_T = C_{T-PTCH} - C_{T-PGK}$. *PTCH* levels in tumours were presented as a ratio to levels detected in adjacent normal tissue (Fig. 3a).

Hh-responsive reporter assays

Hh-responsive firefly luciferase and control SV40 Renilla luciferase reporter assays were performed on subconfluent triplicate cultures as described previously²³. Two days after transfection, culture medium was replaced for a 2-day culture period with assay medium: RPMI-1640 (Bio-Whittaker) supplemented with 0.5% (established cell lines) or 20% (first-passage xenografts) FBS and containing combinations of 5E1 anti-Hh monoclonal antibody, recombinant doubly lipid-modified Sonic hedgehog (ShhNp) peptide¹³, cyclopamine purified from *Veratrum* extract or tomatidine (ICN Pharmaceuticals) at the concentrations indicated in the main text. Lysates were prepared and analysed as described elsewhere¹³.

Proliferation assays

Cells were cultured in triplicate in 96-well plates in assay media to which 5E1 monoclonal antibody, ShhNp and/or cyclopamine were added at 0 h at concentrations indicated in the main text. Viable cell mass was determined by optical density measurements at 490 nm (OD_{490}) at 2 and 4 days using the CellTiter96 (Promega) colorimetric assay. Relative growth was calculated as $OD(\text{day } 4) - OD(\text{day } 2)/OD(\text{day } 2)$.

Xenograft treatment

In accordance with approved Johns Hopkins University Animal Care and Use Committee protocols, HUCCT1 tumours ($n = 18$) were grown in athymic (nude) mice and treated with cyclopamine ($50 \text{ mg kg}^{-1} \text{ d}^{-1}$, subcutaneous injection) or control vehicle as described previously¹⁵.

Received 28 May; accepted 25 July 2003; doi:10.1038/nature01972.

Published online 14 September 2003.

- Bale, A. E. & Yu, K. P. The hedgehog pathway and basal cell carcinomas. *Hum. Mol. Genet.* **10**, 757–762 (2001).
- Wechsler-Reya, R. & Scott, M. P. The developmental biology of brain tumors. *Annu. Rev. Neurosci.* **24**, 385–428 (2001).
- Taipale, J., Cooper, M. K., Maiti, T. & Beachy, P. A. Patched acts catalytically to suppress the activity of Smoothened. *Nature* **418**, 892–897 (2002).
- Taipale, J. & Beachy, P. A. The Hedgehog and Wnt signalling pathways in cancer. *Nature* **411**, 349–354 (2001).
- Ingham, P. W. & McMahon, A. P. Hedgehog signaling in animal development: paradigms and principles. *Genes Dev.* **15**, 3059–3087 (2001).
- Chen, J. K., Taipale, J., Cooper, M. K. & Beachy, P. A. Inhibition of Hedgehog signaling by direct binding of cyclopamine to Smoothened. *Genes Dev.* **16**, 2743–2748 (2002).
- Watkins, D. N. *et al.* Hedgehog signalling within airway epithelial progenitors and in small-cell lung cancer. *Nature* **422**, 313–317 (2003).
- Roberts, D. J., Smith, D. M., Goff, D. J. & Tabin, C. J. Epithelial–mesenchymal signaling during the regionalization of the chick gut. *Development* **125**, 2791–2801 (1998).
- van den Brink, G. R. *et al.* Sonic hedgehog regulates gastric gland morphogenesis in man and mouse. *Gastroenterology* **121**, 317–328 (2001).
- Ramalho-Santos, M., Melton, D. A. & McMahon, A. P. Hedgehog signals regulate multiple aspects of gastrointestinal development. *Development* **127**, 2763–2772 (2000).
- Hebrok, M. Hedgehog signaling in pancreas development. *Mech. Dev.* **120**, 45–57 (2003).
- Bitgood, M. J. & McMahon, A. P. Hedgehog and Bmp genes are coexpressed at many diverse sites of cell–cell interaction in the mouse embryo. *Dev. Biol.* **172**, 126–138 (1995).
- Taipale, J. *et al.* Effects of oncogenic mutations in Smoothened and Patched can be reversed by cyclopamine. *Nature* **406**, 1005–1009 (2000).

14. Cooper, M. K., Porter, J. A., Young, K. E. & Beachy, P. A. Plant-derived and synthetic teratogens inhibit the ability of target tissues to respond to Sonic hedgehog signaling. *Science* **280**, 1603–1607 (1998).
15. Berman, D. M. *et al.* Medulloblastoma growth inhibition by hedgehog pathway blockade. *Science* **297**, 1559–1561 (2002).
16. Ericson, J., Morton, S., Kawakami, A., Roelink, H. & Jessell, T. M. Two critical periods of Sonic Hedgehog signaling required for the specification of motor neuron identity. *Cell* **87**, 661–673 (1996).
17. Wang, L. C. *et al.* Conditional disruption of hedgehog signaling pathway defines its critical role in hair development and regeneration. *J. Invest. Dermatol.* **114**, 901–908 (2000).
18. Fuse, N. *et al.* Sonic hedgehog protein signals not as a hydrolytic enzyme but as an apparent ligand for patched. *Proc. Natl Acad. Sci. USA* **96**, 10992–10999 (1999).
19. Goodrich, L. V., Milenkovic, L., Higgins, K. M. & Scott, M. P. Altered neural cell fates and medulloblastoma in mouse patched mutants. *Science* **277**, 1109–1113 (1997).
20. Chen, X. & Yang, C. S. Esophageal adenocarcinoma: a review and perspectives on the mechanism of carcinogenesis and chemoprevention. *Carcinogenesis* **22**, 1119–1129 (2001).
21. Peek, R. M. Jr *Helicobacter pylori* strain-specific modulation of gastric mucosal cellular turnover: implications for carcinogenesis. *J. Gastroenterol.* **37** (Suppl. 13), 10–16 (2002).
22. Lowenfels, A. B. & Maisonneuve, P. Epidemiologic and etiologic factors of pancreatic cancer. *Hematol. Oncol. Clin. North Am.* **16**, 1–16 (2002).
23. Chen, J. K., Taipale, J., Young, K. E., Maiti, T. & Beachy, P. A. Small molecule modulation of Smoothened activity. *Proc. Natl Acad. Sci. USA* **99**, 14071–14076 (2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank E. Traband and K. Young for technical assistance, E. Montgomery, K. Miyazaki and J. Harmon for cell lines, J. Chen for help with cyclopamine purification and P. Fussell for help with figures. This work was supported by the family of Margaret Lee and grants from the National Institutes of Health (D.M.B., A.M., J.R.E. and P.A.B.). P.A.B. is an investigator and M.R.G. a medical fellow of the Howard Hughes Medical Institute.

Competing interests statement The authors declare competing financial interests: details accompany the paper on www.nature.com/nature.

Correspondence and requests for materials should be addressed to P.A.B. (pbeachy@jhmi.edu).

Hedgehog is an early and late mediator of pancreatic cancer tumorigenesis

Sarah P. Thayer^{1*}, Marina Pasca di Magliano^{4*}, Patrick W. Heiser⁴, Corinne M. Nielsen^{1,2}, Drucilla J. Roberts², Gregory Y. Lauwers², Yan Ping Qi⁴, Stephan Gysin⁵, Carlos Fernández-del Castillo¹, Vijay Yajnik³, Bozena Antoniu¹, Martin McMahon⁵, Andrew L. Warshaw¹ & Matthias Hebrok⁴

¹Department of Surgery, ²Department of Pathology, and ³Gastroenterology Unit, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114, USA

⁴Diabetes Center, Department of Medicine, and ⁵Cancer Research Institute, University of California, San Francisco, California 94143, USA

* These authors contributed equally to this work

Hedgehog signalling—an essential pathway during embryonic pancreatic development, the misregulation of which has been implicated in several forms of cancer—may also be an important mediator in human pancreatic carcinoma^{1–8}. Here we report that sonic hedgehog, a secreted hedgehog ligand, is abnormally expressed in pancreatic adenocarcinoma and its precursor lesions: pancreatic intraepithelial neoplasia (PanIN). Pancreata of Pdx–Shh mice (in which Shh is misexpressed in the pancreatic endoderm) develop abnormal tubular structures, a phenocopy of human PanIN-1 and -2. Moreover, these PanIN-like lesions also contain mutations in K-ras and overexpress HER-2/*neu*, which are genetic mutations found early in the progression of human pancreatic cancer. Furthermore, hedgehog signalling remains active in cell lines established from primary and metastatic pancreatic adenocarcinomas. Notably, inhibition of hedgehog

signalling by cyclopamine induced apoptosis and blocked proliferation in a subset of the pancreatic cancer cell lines both *in vitro* and *in vivo*. These data suggest that this pathway may have an early and critical role in the genesis of this cancer, and that maintenance of hedgehog signalling is important for aberrant proliferation and tumorigenesis.

Sonic hedgehog (SHH) is misexpressed in human adenocarcinoma and its precursor lesions. SHH expression was determined using *in situ* hybridization to detect SHH messenger RNA and immunohistochemistry (IHC) to detect the protein with an antibody directed against SHH⁹. Pancreatic tissues were obtained from 20 specimens resected for pancreatic cancer. Control pancreatic tissues with no evidence of abnormality or autolysis upon histological evaluation were obtained from autopsy specimens or from pancreatic resections for trauma. In normal adult human pancreata, no SHH was detected in the islets, acini or ductal epithelium (Fig. 1a). However, evaluation of pancreata from patients with adenocarcinoma reveals that SHH is aberrantly expressed in 70% of specimens. Normal ductal epithelium does not express detectable levels of SHH (Fig. 1b); however, as the ductal epithelium shows increasing degrees of atypia, PanIN-1 to -3 (Fig. 1c–e), a higher expression of SHH is observed. SHH expression is also detected in the malignant epithelium of adenocarcinoma samples (Fig. 1f). This expression pattern was also confirmed by our *in situ* hybridization for SHH mRNA (Supplementary Fig. 1).

Loss of regulation in this pathway has been implicated in several human cancers^{10,11}. Thus in order to determine the potential role of SHH misexpression in the adult human pancreas, pancreata from

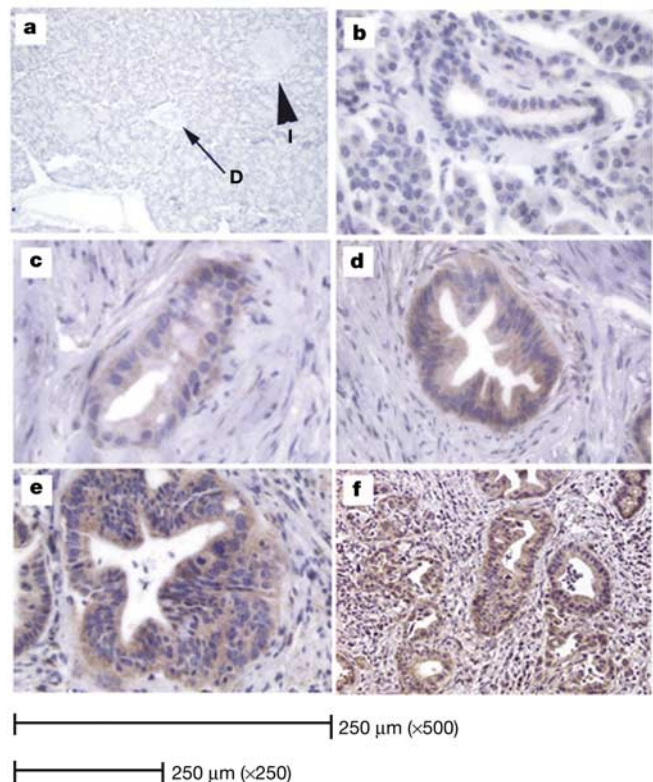


Figure 1 Immunohistochemical identification of SHH. **a**, Normal human pancreas. No specific staining for SHH protein was identified in acini, islets (I) or ducts (D) (×125 magnification). **b**, No SHH expression is detected in normal ductal epithelium (×500 magnification). **c**, PanIN-1 expresses minimal amounts of SHH (×500). **d**, PanIN-2 expresses moderate levels of SHH (×250). **e**, PanIN-3 (×250) and **f**, invasive cancer (×125): moderate to high levels of SHH.

transgenic mice (gift of H. Edlund) in which *Shh* misexpression was driven by the pancreatic-specific *Pdx-1* promoter were histologically and immunohistochemically analysed.

A total of four pancreata from three-week-old *Pdx-Shh* mice

were histologically evaluated by a gastrointestinal pathologist (G.Y.L.). All four expressed a significant intestinal phenotype of the pancreatic epithelium. All four also exhibited abnormal tubular complex formations of the epithelium resembling pancreatic cancer precursor lesions (PanINs). These abnormal epithelial changes were characterized and classified according to the pathological classification scheme for human ductal lesions of the pancreas¹².

The transgenic pancreata display a mixed phenotype characterized by the transformation of acini into glandular structures lined by tall intestinal-type epithelium embedded in cellular spindle stroma. Focally, the metaplastic epithelium is composed of columnar cells with basally located nuclei and abundant supranuclear mucin (Fig. 2a). This phenotype is similar to human PanIN-1 (Fig. 2b), an early lesion in the stepwise sequence of pancreatic carcinogenesis. Furthermore, some glands are characterized by a lesser degree of mucin production but a higher grade of atypia, with nuclear disarray and stratification, mitosis and apoptosis (Fig. 2c), representing a higher degree of dysplastic change, similar to human PanIN-2 (Fig. 2d). According to the tumour progression model of

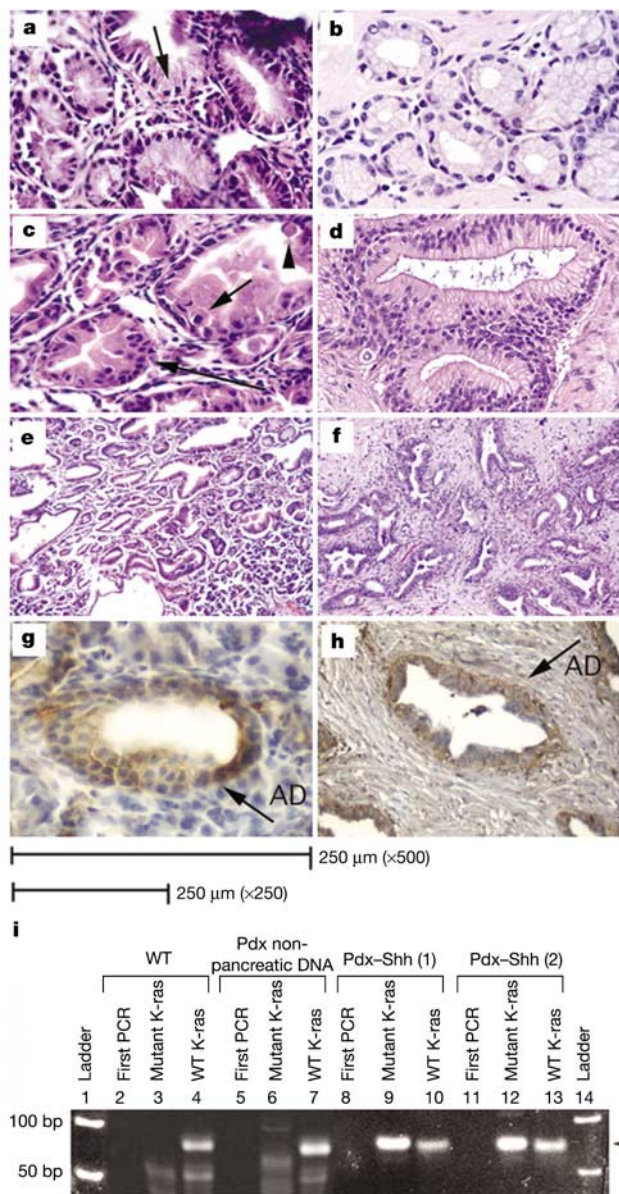


Figure 2 Histological comparison of human PanINs and *Pdx-Shh* mouse pancreata. **a**, *Pdx-Shh* pancreata display abnormal intestinal epithelial changes (arrow identifies abundant supranuclear mucin) that resemble human PanIN-1 (**b**) ($\times 500$). **c**, *Pdx-Shh* pancreata ($\times 500$) also exhibit a higher degree of atypia, with nuclear disarray and stratification (long arrow), mitosis (short arrow) and apoptosis (arrowhead), resembling human PanIN-2 (**d**) ($\times 250$). **e**, **f**, Low-magnification ($\times 125$) view of *Pdx-Shh* pancreata (**e**) and tubular complexes seen in human adenocarcinoma (**f**). **g**, *Pdx-Shh* mice ($\times 500$) and human PanINs (**h**) ($\times 250$) overexpress HER-2/*neu*, identified by a brown stain (AD; abnormal ductal epithelium). **i**, Mutant-allele-specific PCR amplification using mismatched primers (mutant K-ras) to detect codon 12 K-ras mutation. Wild-type primer (WT K-ras) is used as control. Lanes 1 and 14, ladder; 2, 5, 8, 11, first round of PCR (first PCR) reveals no detectable band, as expected. Wild-type mice express only wild-type K-ras (lanes 3, 4). Abnormal pancreatic epithelium of two *Pdx-Shh* mice expresses a TGT codon 12 mutation of K-ras (lanes 9, 12) and wild-type K-ras (lanes 10, 13), whereas the extrapancreatic tissues of *Pdx-Shh* mice outside the *Shh* expression domain express only wild-type K-ras (lanes 6, 7). The arrowhead identifies the 70-bp amplicon.

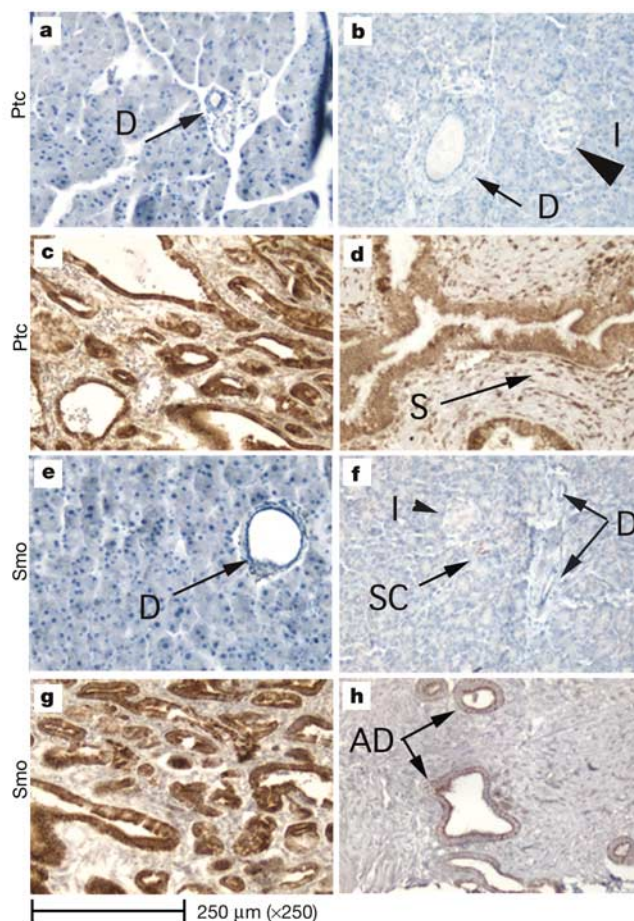


Figure 3 Hedgehog signalling pathway. **a-d**, Immunohistochemical identification of Ptc1 receptor. The Ptc1 receptor is not identified in normal acini, ducts or islets in either mouse (**a**) or human (**b**) pancreata. Ptc1 (brown stain) is markedly overexpressed in the abnormal pancreatic epithelium of both *Pdx-Shh* mice (**c**) and neoplastic human pancreata (**d**). In humans, PTC1 is also overexpressed throughout the reactive stroma (S) that surrounds the epithelium. **e-h**, Immunohistochemical identification of Smoothened (Smo). Smoothened is not detected in normal duct epithelium (D) of normal mouse (**e**) or human (**f**) pancreata. Very infrequently Smo is detected in scattered cells (SC) in human pancreata; however, Smo (brown stain) is detected in the abnormal epithelium of *Pdx-Shh* mice (**g**) and neoplastic human pancreata (**h**) (arrows). The magnification is $\times 250$ for all images.

pancreatic carcinoma, these features are suggestive of an increasing neoplastic potential.

The histological progression model of pancreatic adenocarcinoma is also associated with progressive genetic alterations¹³. The transgenic mice pancreata, which histologically resemble PanIN-1 and -2, also express some of the early genetic alterations observed in the human progression model. Three pancreata overexpress HER-2/*neu* in the epithelium (Fig. 2g). In addition, preliminary data indicate that a codon 12 mutation in K-ras was present in two pancreata (Fig. 2i): changes noted to occur in human PanIN-1 and -2 lesions^{14,15}. Thus the Pdx-Shh mice develop abnormal pancreata with morphological and genetic changes that resemble human ductal PanIN.

Hedgehog (Hh) signalling is activated through binding of Shh to a membrane receptor, Ptc. This ligand-receptor interaction inhibits Ptc function, thereby allowing the de-repression of another transmembrane protein, smoothened (Smo), which ultimately leads to the activation of the Hh pathway and to the expression of Hh-specific genes, one of which is Ptc1 (ref. 16). Ptc1 expression (determined by IHC) is not detected in normal mouse or human pancreata (Fig. 3a, b). However, Ptc1 protein expression was detected throughout the abnormal epithelium in both Pdx-Shh mice (Fig. 3c) and human neoplastic pancreata (Fig. 3d). Furthermore, in human neoplastic pancreata, PTC1 was also detected in the

reactive mesenchymal cells that surround the abnormal epithelium (Fig. 3d).

Another member of the Hh pathway that is similarly overexpressed in both human neoplastic and mouse transgenic pancreata is Smo. In normal human and mouse pancreata Smo is not generally detected by IHC in the ducts or acini. Very infrequently Smo can be detected in an occasional duct cell or scattered acinar cell within the normal pancreas (Fig. 3f). However, Smo was overexpressed in the abnormal pancreatic epithelium found in Pdx-Shh mice (Fig. 3g), as well as in the epithelium of human adenocarcinoma and precursor lesions (Fig. 3h). The identification of Hh signalling members overexpressed in both mesenchyme and abnormal epithelium suggests that misexpression of Shh may be acting through both a paracrine and an autocrine mechanism to cause many of the morphological changes that characterize pancreatic cancer. To determine whether the HH pathway has an equally important role in the malignant biological behaviour of this tumour, we undertook work with pancreatic cancer cell lines.

Twenty-six human adenocarcinoma lines were screened for expression of HH signalling components. These cell lines were derived either from primary tumours (for example, the Panc series Panc 01.28 to 10.05 (ref. 17)) or from liver, lymph node and spleen metastases (CFPAC1, Hs 766T and SW 1990, respectively). All lines tested expressed two or more components of HH signalling,

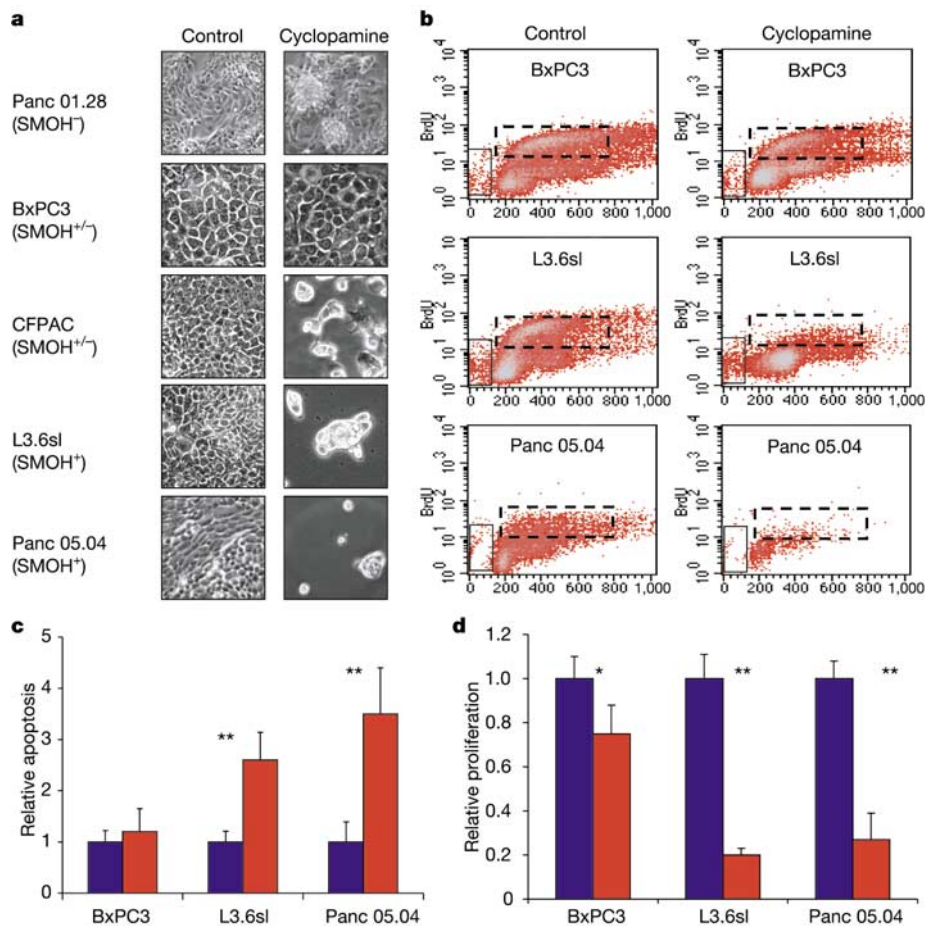


Figure 4 Effects of cyclopamine treatment on pancreatic adenocarcinoma cells.

a, Cell cultures of five pancreatic adenocarcinoma cell lines untreated (control) or treated with 10 μ M cyclopamine for 7 days. **b**, Representative FACS histograms of three cell lines after BrdU incorporation ($n = 4$). The x axes show DNA content whereas the y axes show BrdU level; vertical boxes mark apoptotic cells; dashed horizontal boxes mark proliferating cells. **c**, Quantification of apoptotic cells measured in **b** ($n = 4$). To compare

treated and untreated cells, the number of apoptotic cells in control samples was adjusted to 1. **d**, Quantification of proliferating cells measured in **b** ($n = 4$). To compare treated and untreated cells, the number of proliferating cells in control samples was adjusted to 1. Error bars indicate standard deviation. Asterisk, $P < 0.05$; double asterisks, $P < 0.01$. SMO^{H-}, no expression; SMO^{H+/+}, low level expression; SMO^{H+}, strong expression.

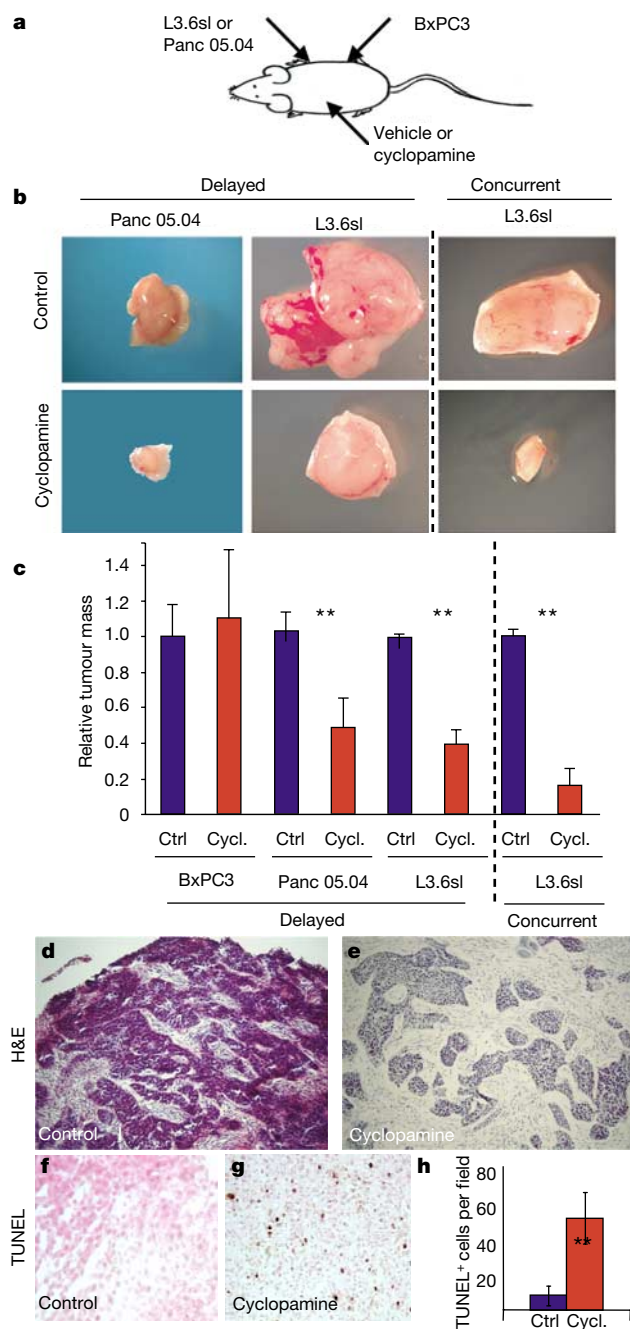


Figure 5 Cyclopamine treatment blocks tumour formation of human pancreatic adenocarcinoma cells after transplantation into nude mice. **a**, Schematic indicating sites of tumour cell and cyclopamine/vehicle injections. **b**, Isolated tumours derived from control or cyclopamine-treated L3.6sl and Panc 05.04 adenocarcinoma cells. Cyclopamine/vehicle injections were initiated either after palpable tumours had formed (delayed) or simultaneously with injection of tumour cells (concurrent). All pictures are shown at the same magnification. **c**, Weight of isolated tumours. Untreated control tumours of each cell line were adjusted to 1 to allow comparison of relative change in tumour mass. For delayed cyclopamine/vehicle injections, the values are: BxPC3, control 1 ($n = 6$), cyclopamine 1.1 ($n = 5$); Panc 05.04, control 1 ($n = 5$), cyclopamine 0.48 ($n = 4$); L3.6sl, control 1 ($n = 5$), cyclopamine 0.39 ($n = 4$). The values for concurrent cyclopamine/vehicle injections are: L3.6sl, control 1 ($n = 4$), cyclopamine 0.16 ($n = 4$). Error bars indicate standard deviation. Double asterisks, $P < 0.01$. **d–h**, Histological analysis of the effect of cyclopamine treatment on L3.6sl-derived tumours. **d**, **e**, Haematoxylin/eosin staining of sections through the peripheral tumour regions. **f**, **g**, TUNEL staining of apoptotic cells in control (**f**) and cyclopamine-treated (**g**) tumours. **h**, Quantification of TUNEL-positive cells in control (blue) and cyclopamine-treated (red) tumours. Error bars indicate s.e.m. Double asterisks, $P < 0.01$.

including *PTCH1*, *SMO*, *HIP* and *GLI1* (Supplementary Fig. 2). Thus, sustained HH signalling activity is detected in pancreatic adenocarcinoma cell lines isolated from both primary and metastatic tumours.

Cyclopamine, a steroidal alkaloid that inhibits Hh signalling through direct interaction with Smo^{18–20}, was used to test whether pancreatic adenocarcinoma cells require HH signalling for proliferation and survival. Initially five pancreatic cancer cell lines—BxPC3, Panc 01.28, CFPAC, L3.6sl and Panc 05.04—were qualitatively assessed. After 7 days of treatment, cyclopamine produced no change in cell density or morphology in control samples that included either untreated cells or cells treated with tomatidine, an inactive cyclopamine analogue. By contrast, a marked reduction in cell density was noted in cell lines CFPAC, L3.6sl and Panc 05.04. Morphologically, the majority of these cells detached from tissue culture plates during the treatment period, and the few remaining cells appeared aplastic (Fig. 4a, and data not shown).

Three representative pancreatic cancer cell lines—BxPC3-*SMO*^{low}, L3.6sl-*SMO*^{high} and Panc 05.04-*SMO*^{high}—were then used to quantitatively determine the effects of cyclopamine on proliferation and apoptosis by means of FACS analysis after 5-bromodeoxyuridine (BrdU) labelling. Cyclopamine did not induce apoptosis and only marginally reduced proliferation in the BxPC3-*SMO*^{low} control cell line. By contrast, apoptosis increased 2.5- to 3.5-fold accompanied by a 75–80% reduction in proliferation in the *SMO*^{high} cyclopamine-responsive cell lines L3.6sl and Panc 05.04 (Fig. 4b–d).

Our initial studies with five cell lines showed that only a subset of pancreatic cancer cell lines are susceptible to cyclopamine treatment. This could be the result of activating mutations of the HH pathway downstream of SMO, which should not be inhibited by this antagonist. To determine the percentage of cancer cell lines susceptible to cyclopamine treatment, we expanded our analysis to a total of 13 lines. Twelve of thirteen lines analysed expressed SMO; half of these (50%) responded to cyclopamine treatment. As predicted, growth of *SMO*-negative Panc 01.28 cells was not affected (Supplementary Fig. 3). Thus our data suggest that mutations in the HH pathway downstream of SMO might contribute to the aberrant proliferation of these cells. However, we cannot exclude the possibility that mutations in other, HH-independent, pathways regulate proliferation in cyclopamine-resistant pancreatic cancer cell lines.

To demonstrate that cyclopamine-induced inhibition of cancer cell proliferation is secondary to inactivation of the HH pathway, a Gli-luciferase reporter construct containing Gli-binding sites upstream of a TK minimal promoter²¹ was transfected into control and cyclopamine-responsive cell lines. GLI1 is a downstream transcription factor and transcriptional target of the Hh pathway²². All cell lines tested expressed GLI1 (Supplementary Fig. 2), and transfection of the Gli-reporter construct resulted in significant elevation of luciferase activity (Supplementary Fig. 4a), further indicating that Hh signalling is active in adenocarcinoma cells. Treatment of transfected cells with cyclopamine abolishes luciferase activity in responsive cell lines, whereas control cells display only an insignificant reduction in luciferase activity. Active HH signalling thus appears to be essential for cell survival and proliferation in a subset of cultured human pancreatic cancer cell lines.

As a more definitive test of the ability of cyclopamine to inhibit tumour cell proliferation, control (BxPC3) and cyclopamine-responsive cells (L3.6sl and Panc 05.04) were injected subcutaneously into immunocompromised nude mice (Fig. 5a). Two treatment models were used—concurrent or delayed treatments—in which cyclopamine was given either at the time of tumour implantation (concurrent) or when the subcutaneous tumours were palpable (delayed). Cyclopamine or carrier control was given for 7

days, at which time the tumours were excised and weighed. In the delayed treatment model, no difference in weight was noted between control and cyclopamine-treated BxPC3-SMO^{low} tumours (Fig. 5c). By contrast, a 50–60% decrease in tumour mass was observed in Panc 05.04- and L3.6sl-derived tumours, respectively (Fig. 5b, c)—an even more marked effect was noted in the concurrent treatment model, which revealed an 84% reduction in tumour mass of L3.6sl-derived tumours (Fig. 5b, c).

Histological analysis of L3.6sl-derived tumours (delayed treatment) revealed a compact mass of epithelial cells in untreated controls, whereas treated tumours appeared as loose epithelial cell aggregates with a larger amount of interspersed mesenchymal cells (Fig. 5d, e). TdT-mediated dUTP nick end labelling (TUNEL) assays demonstrated that cyclopamine treatment resulted in a sixfold increase in apoptotic cells compared with untreated controls (Fig. 5f–h). Therefore both *in vitro* and *in vivo* experiments reveal that HH signalling is important for proliferation and survival and thus may be an important mediator of the malignant biological behaviour of pancreatic adenocarcinomas in humans.

The observation that HH signalling has a critical and central role in pancreatic development and results in malignant transformation when mutated prompted us to ask about its potential role in pancreatic cancer^{16,23–25}. Our experiments lead us to believe that misregulation of HH signalling has a critical early role in the initiation and a later role in the maintenance of human pancreatic ductal adenocarcinoma, a hypothesis that is supported by findings in the accompanying paper³¹. Our evidence is based on comparative analysis of pancreata from Pdx1–Shh transgenic mice and of human adenocarcinoma. We find that HH signalling components are undetectable in normal human ductal epithelium but are strongly expressed in pancreatic precursor (PanIN) and invasive lesions. When Shh is misexpressed in Pdx–Shh mice, we detected the presence of abnormal pancreata with features that histologically, immunohistochemically and genetically resemble human PanIN lesions. Moreover, our analysis of human pancreatic cancer cell lines indicates that the HH signalling pathway is required for both cell division and the suppression of apoptosis. An important but unresolved question concerns the identity of the pancreatic cells that become transformed in response to increased HH signalling. Although pancreatic stem cells have not yet been isolated, results from studies in other tissues suggest that increased HH signalling in progenitor cells can elicit cancer formation. Future studies will be needed to elucidate the role of HH activity in stem-cell-mediated tissue regeneration and pancreatic adenocarcinomas.

Pancreatic cancer, the fourth highest cause of death from cancer in the United States, remains an incurable and rapidly lethal disease, with a 5-yr survival rate of less than 3%. Many factors contribute to this poor prognosis, but ultimately it is our lack of knowledge of the molecular determinants involved in the pathogenesis and progression of this tumour that has limited our ability to design effective treatments. Identification of a role for the HH pathway in the initiation and maintenance of pancreatic cancer suggests that this pathway may hold promise for new diagnostic and therapeutic approaches. □

Methods

Tissue preparation

Human tissue specimens were collected in accordance with IRB approval from the Massachusetts General Hospital. Archived specimens were fixed with formalin and paraffin-embedded according to standard institutional protocol. Wild type and Pdx–Shh mouse tissues were fixed in 4% paraformaldehyde, then paraffin-embedded as described above. Tissues were cut into 4–6-µm sections and applied to charged ProbeOn Plus slides (Fisher).

Transgenic mice

Initial pancreatic tissue (gift of H. Edlund) was in a B6/CBA background. Transient transgenic mice (founder mice) were generated by pronuclear injection of a *Not1/BamH1*

expression cassette containing a 4.5-kilobase (kb) *Not1–NaeI* genomic fragment of the Pdx promoter cloned in front of a 2.6-kb *XhoI* fragment of full-length rat *Shh* complementary DNA (vector also a gift of H. Edlund) into a B6/C3F1 background as previously described²⁶.

Immunohistochemistry

Single-antibody detection was accomplished as previously described²⁶, with the following protocol modifications: endogenous peroxidase activity was blocked using 3% H₂O₂ in methanol for 10 min at room temperature. Antigen retrieval was achieved by boiling tissue in 0.01 M sodium citrate, pH 2.0, 6.0, or 8.0 for 10 min. Tissues were blocked first with 10% NGS in TBS for 40 min at room temperature, then with streptavidin followed by biotin for 20 min each at 37°C. All primary antibodies were incubated overnight at 4°C. Primary antibodies used were: Shh, Ptc1 and Smo (1:250; Santa Cruz). Secondary biotinylated antibodies (Vector Laboratories) were applied in 1:500 dilution for 1 h at room temperature. Detection of protein was visualized by brown pigmentation via standard DAB protocol. Slides were lightly counterstained with haematoxylin.

Detection of K-ras codon 12 mutation

The mutant-allele-specific amplification method was used for detection of K-ras codon 12 mutation as previously described^{27,28}. DNA was extracted from microdissected pancreata. Primers used for first-round PCR amplification for K-ras were K12 forward primer 5'-CGCGGGCGCTGAATGACTGA-3' and K12 reverse primer 5'-TCGTAGGGTCATACTCATCC-3'. PCR conditions were: 94°C for 1 min, 54°C for 1 min, and 72°C for 0.5 min for 20 cycles. Then a second PCR was performed on 2.0 µl of the first reaction for an additional 40 cycles under similar conditions, using as the new upstream primer either a mismatched primer to amplify the specific mutant band (GGT to TGT), or a wild-type primer, producing a 72-base-pair (bp) mutated or 71-bp wild-type fragment. The second PCR downstream primer sequence is K12 R nest in 5'-CCACAAAGTGATTCTGAATTA-3'. Mismatched upstream primer sequence is mutant K12 (TGT) 5'-TTGTGGTGGTTGGAGCTT-3'; and wild-type sequence is WT K12 5'-TGTGGTGGTTGGAGCTGG-3'. Amplifications were run on a 10% TBE polyacrylamide gel and stained with SYBR green (Molecular Probes).

Cell culture

Human pancreatic adenocarcinoma cell lines HPAC, SW1990, Mpanc-96, SU86.86, PL45, Panc 10.05, Panc 8.13 and Panc 2.03 were obtained from the American Tissue Culture Collection; cell lines MiaPaCa2, Panc-1, CFPAC1, HPAC1, Capan-2, AsPC1, Hs766T and BxPC3 were a gift from Schering Plough. The cell lines COLO357, L3.3, L3.6sl and L3.6pl were a gift from I. Fidler; cell lines Panc 3.07, Panc 5.04, Panc 2.13, Panc 6.03, Panc 4.21 and Panc 1.28 were a gift from E. Jaffee. BxPC3 and all the Panc cell lines were grown in RPMI medium (Gibco) supplemented with 10% fetal bovine serum (Gibco), L-glutamine and penicillin/streptomycin; medium for Panc cell lines was also supplemented with insulin–transferrin–selenium (Gibco). The CFPAC, Panc1, L3.6sl and L3.6pl cell lines were grown in DMEM without phenol red (CellGro), supplemented with 10% fetal calf serum. To test for cyclopamine responsiveness, cells were grown for 7 days in control medium containing tomatidine (Sigma) or DMSO alone or experimental medium containing cyclopamine (10 µM, Toronto Research Chemicals; the cyclopamine dose–response is shown in Supplementary Fig. 4b). We changed the medium every 2 days. Pictures showing cell morphology were taken with a Nikon Eclipse TE300.

BrdU incorporation assay

Cells were grown for 3 or 4 days in medium containing tomatidine (control, Sigma) or cyclopamine (10 µM, Toronto Research Chemicals). The medium was changed every 48 h. Cells were pulsed with 10 µM BrdU during the final 2 h of culture. BrdU was detected with a fluorescein isothiocyanate-conjugated anti-BrdU antibody (BD Biosciences); total DNA was stained with 7-AAD. FACS analysis was performed according to the BD Biosciences BrdU flow kit instruction manual. Cells in S phase were defined as a cell population that had incorporated BrdU, with DNA content comprising between 2N and 4N. According to the manual, apoptotic cells were defined as a subpopulation of G0/G1 cells with DNA content lower than the diploid amount.

Allograft treatment *in vivo*

Allograft treatment *in vivo* was performed according to ref. 19 with minor modifications. A total of 0.1 ml Hanks' balanced salt solution and matrigel (1:1) containing 2 × 10⁶ cells was injected subcutaneously into CD-1 nude mice. Tumours were grown for 4 days to a minimum volume of 125 mm³; treatment was initiated simultaneously for all subjects. Mice were injected subcutaneously with vector alone (trioline:ethanol 4:1 v/v) or a cyclopamine suspension (1.2 mg per mouse in trioline:ethanol 4:1 v/v) daily for 7 days. At the end of the treatment period, tumours were excised from mice, weighed and then fixed for 3 h at 4°C with 4% paraformaldehyde, embedded in paraffin wax and sectioned (6 µm). Apoptotic cells were identified by TUNEL using recombinant Tdt (Invitrogen) as previously described²⁹. Sections were then counterstained with eosin. Eight ×20-magnified fields from regions corresponding to the exterior, middle and interior of two control and two cyclopamine-treated tumours were chosen at random. We counted the number of TUNEL-positive nuclei manually. Haematoxylin/eosin staining was done as previously described³⁰.

Received 30 May; accepted 28 August 2003; doi:10.1038/nature02009.

Published online 14 September 2003.

1. Capdevila, J., Estrada, M. P., Sanchez-Herrero, E. & Guerrero, I. The *Drosophila* segment polarity gene patched interacts with decapentaplegic in wing development. *EMBO J.* **13**, 71–82 (1994).

2. Hardcastle, Z., Mo, R., Hui, C.-C. & Sharpe, P. T. The Shh signalling pathway in tooth development: defects in Gli2 and Gli3 mutants. *Development* **125**, 2803–2811 (1998).
3. Marigo, V. & Tabin, C. Regulation of patched by sonic hedgehog in the developing neural tube. *Proc. Natl Acad. Sci. USA* **93**, 9346–9351 (1996).
4. Riddle, R. D., Johnson, R. L., Lauffer, E. & Tabin, C. Sonic hedgehog mediates the polarizing activity of the ZPA. *Cell* **75**, 1401–1416 (1993).
5. St-Jacques, B. *et al.* Sonic hedgehog signaling is essential for hair development. *Curr. Biol.* **8**, 1058–1068 (1998).
6. Roberts, D. J. in *Development of the Gastrointestinal Tract* (eds Sanderson, I. R. & Walker, W. A.) 1–12 (Decker, Hamilton, Ontario, 2000).
7. Hebrok, M. Hedgehog signaling in pancreas development. *Mech. Dev.* **120**, 45–47 (2003).
8. Watkins, D. N. *et al.* Hedgehog signaling within airway epithelial progenitors and in small-cell lung cancer. *Nature* **422**, 313–317 (2003).
9. van den Brink, G. R. *et al.* Sonic hedgehog regulates gastric gland morphogenesis in man and mouse. *Gastroenterology* **121**, 317–328 (2001).
10. Johnson, R. L. *et al.* Human homolog of patched, a candidate gene for the basal cell nevus syndrome. *Science* **272**, 1668–1671 (1996).
11. Wicking, C., Smyth, I. & Bale, A. The hedgehog signalling pathway in tumorigenesis and development. *Oncogene* **18**, 7844–7851 (1999).
12. Hruban, R. *et al.* Pancreatic intraepithelial neoplasia. *Am. J. Surg. Pathol.* **25**, 579–586 (2001).
13. Wilentz, R. E. *et al.* Loss of expression of Dpc4 in pancreatic epithelial neoplasia: evidence that DPC4 inactivation occurs late in neoplastic progression. *Cancer Res.* **60**, 2002–2006 (2000).
14. Day, J. D. *et al.* Immunohistochemical evaluation of HER-2/neu oncogene expression in pancreatic adenocarcinoma and pancreatic intraepithelial neoplasms. *Hum. Pathol.* **27**, 119–124 (1996).
15. Wilentz, R. E. *et al.* Inactivation of the p16 (*INK4A*) tumor-suppressor gene in pancreatic duct lesions: loss of intranuclear expression. *Cancer Res.* **58**, 4740–4744 (1998).
16. Hebrok, M., Kim, S. K., St-Jacques, B., McMahon, A. P. & Melton, D. A. Regulation of pancreas development by hedgehog signaling. *Development* **127**, 4905–4913 (2000).
17. Jaffee, E. M. *et al.* Development and characterization of a cytokine-secreting pancreatic adenocarcinoma vaccine from primary tumors for use in clinical trials. *Cancer J. Sci. Am.* **4**, 194–203 (1998).
18. Taipale, J. *et al.* Effects of oncogenic mutations in Smoothened and Patched can be reversed by cyclopamine. *Nature* **406**, 1005–1009 (2000).
19. Berman, D. M. *et al.* Medulloblastoma growth inhibition by hedgehog pathway blockade. *Science* **297**, 1559–1561 (2002).
20. Incardona, J. P., Gaffield, W., Kapur, R. P. & Roelink, H. The teratogenic Veratrum alkaloid cyclopamine inhibits sonic hedgehog signal transduction. *Development* **125**, 3553–3562 (1998).
21. Sasaki, H., Hui, C., Nakafuku, M. & Kondoh, H. A binding site for Gli proteins is essential for HNF-3 beta floor plate enhancer activity in transgenics and can respond to Shh *in vitro*. *Development* **124**, 1313–1322 (1997).
22. Lee, J., Platt, K. A., Censullo, P. & Ruiz I Altaba, A. Gli1 is a target of Sonic hedgehog that induces ventral neural tube development. *Development* **124**, 2537–2552 (1997).
23. Hebrok, M., Kim, S. K. & Melton, D. A. Notochord repression of endodermal Sonic hedgehog permits pancreas development. *Genes Dev.* **12**, 1705–1713 (1998).
24. Roberts, D. J., Smith, D. M., Goff, D. & Tabin, C. Epithelial-mesenchymal signaling during the regionalization of the chick gut. *Development* **125**, 2791–2801 (1998).
25. Ramalho-Santos, M., Melton, D. A. & McMahon, A. P. Hedgehog signals regulate multiple aspects of gastrointestinal development. *Development* **127**, 2763–2772 (2000).
26. Apelqvist, A., Ahlgren, U. & Edlund, H. Sonic hedgehog directs specialised mesoderm differentiation in the intestine and pancreas. *Curr. Biol.* **7**, 801–804 (1997).
27. Hasegawa, Y. *et al.* Detection of K-ras mutations in DNAs isolated from feces of patients with colorectal tumors by mutant-allele-specific amplification (MASA). *Oncogene* **10**, 1441–1445 (1995).
28. Ito, Y. *et al.* Frequent detection of K-ras mutation in stool samples of colorectal carcinoma patients after improved DNA extraction: Comparison with tissue samples. *Int. J. Oncol.* **20**, 1263–1268 (2002).
29. Lopez, T. & Hanahan, D. Elevated levels of IGF-1 receptor convey invasive and metastatic capability in a mouse model of pancreatic islet tumorigenesis. *Cancer Cell* **1**, 339–353 (2002).
30. Sander, M. *et al.* Genetic analysis reveals that PAX6 is required for normal transcription of pancreatic hormone genes and islet development. *Genes Dev.* **11**, 1662–1673 (1997).
31. Berman, D. M. *et al.* Widespread requirement for Hedgehog ligand stimulation in growth of digestive tract tumours. *Nature* doi:10.1038/nature01972 (this issue).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements S.P.T. and co-workers are grateful to D. A. Melton and P. A. Donahoe for discussion and review of the manuscript; H. Edlund for supplying the construct and initial pancreatic tissue for the Pdx-Shh transgenic mouse; and C. Tabin for the Shh *in situ* probe. We also thank N. Frost for editorial assistance. This work was supported in part by a grant from the Lustgarten Foundation to S.P.T. M.H. and co-workers thank C. Basbaum, D. Hanahan, I. Herskowitz, T. Kornberg, M. Tempero and members of the Hebrok laboratory for discussions and comments on the manuscript. We thank L. Jaffee for provision of the Panc series of cells. We are grateful to L. Spector for editorial assistance. This work was supported by grants to M.H. from the Juvenile Diabetes Foundation and the National Institutes of Health.

Authors' contributions S.P.T. and M.H. are senior authors.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.P.T. (stayer@partners.org) or M.H. (mhebrok@diabetes.ucsf.edu).

Yip3 catalyses the dissociation of endosomal Rab–GDI complexes

Ulf Sivars, Dikran Aivazian & Suzanne R. Pfeffer

Department of Biochemistry, Stanford University School of Medicine, Stanford, California 94305-5307, USA

Human cells contain more than 60 small G proteins of the Rab family, which are localized to the surfaces of distinct membrane compartments and regulate transport vesicle formation, motility, docking and fusion^{1–3}. Prenylated Rabs also occur in the cytosol bound to GDI^{4,5} (guanine nucleotide dissociation inhibitor), which binds to Rabs in their inactive state. Prenyl Rab–GDI complexes contain all of the information necessary to direct Rab delivery onto distinct membrane compartments^{6–8}. The late endosomal, prenyl Rab9 binds GDI with very high affinity⁹, which led us to propose that there might be a ‘GDI-displacement factor’ to catalyse dissociation of Rab–GDI complexes and to enable transfer of Rabs from GDI onto membranes^{6,10}. Indeed, we have previously shown that endosomal membranes contain a proteinaceous factor that can act in this manner¹⁰. Here we show that the integral membrane protein, Yip3, acts catalytically to dissociate complexes of endosomal Rabs bound to GDI, and to deliver them onto membranes. We propose that the conserved Yip proteins serve as GDI-displacement factors for the targeting of Rab GTPases in eukaryotic cells.

Yip1p is an essential yeast protein that interacts with all yeast Rabs with varying affinities^{11,12}. Cells harbouring a temperature-sensitive allele of *yip1* display an accumulation of endoplasmic reticulum membranes and defects in secretion¹¹. Yeast and human cells contain a family of Yip-related proteins¹³, including Yif1, which interacts with yeast Rabs and Yip1p¹⁴. In mammalian cells, there are at least five Yip-related proteins; they appear to interact with each other and may function as distinct heterodimer pairs. Prenylated Rab acceptor (PRA-1) is the human Yip3 homologue: it interacts with multiple, prenylated Rabs^{15,16}, shows weak interaction with GDI¹⁷ and is localized to the Golgi¹⁸ and late endosomes (see Supplementary Information). The related PRA-2 is present in the endoplasmic reticulum; the carboxy termini of PRA proteins are responsible for their distinct localizations¹⁸.

His-tagged, human Yip3 was expressed in *Escherichia coli* and purified from the membranes after detergent extraction (Fig. 1a). Yip3 is predicted to have four transmembrane domains¹⁹, and the protein was highly susceptible to aggregation if detergent was limiting during purification. With adequate detergent, the protein chromatographed with an apparent mass of a monomer and could be obtained in highly purified form.

GDI binds tightly to Rab9⁹ and blocks release of bound GDP⁴. However, a GDI-displacement factor (GDF) facilitates the generation of free Rab9, which can then exchange bound GDP for ³⁵S-GTPγS (Fig. 1b)¹⁰. This method can detect a GDF, or alternatively, an unusual activity that could act on Rab–GDI complexes and stimulate that Rab’s intrinsic nucleotide exchange rate (a ‘GEF’). Purified, non-tagged, prenyl Rab9–GDI complexes were used to test whether purified human Yip3 could dissociate Rab–GDI complexes as monitored by an increase in the rate of ³⁵S-GTPγS binding. GDI release leads to prenyl-Rab aggregation; assays are carried out with just enough detergent (0.1% CHAPS) to stabilize a released Rab without destabilizing Rab–GDI complexes¹⁰. Once released from GDI, a Rab will exchange bound GDP for GTPγS and then be incapable of rebinding GDI.

Human Yip3 stimulated the rate of nucleotide binding to Rab9 when added to prenyl Rab9–GDI complexes (Fig. 1c). GDF activity

was proportional to the amount of purified protein added, and was abolished upon boiling (Fig. 1c). The reaction displayed linear kinetics for 15 min at 37 °C. Yip3 acted catalytically because as many as ~450 Rab9 molecules were released per Yip3 molecule added. After a 20-min reaction, ~25% of the Rab–GDI complexes were released (Fig. 1c). Addition of more Yip3 after various times of incubation led to complete release of all Rab–GDI complexes (Fig. 1d), suggesting that purified Yip3 loses activity upon incubation at 37 °C; nevertheless, at all times, Yip3 was only present in catalytic amounts. These data demonstrate that the reaction does not reflect simple, stoichiometric binding. In control experiments, Yip3 did not enhance the intrinsic rates of nucleotide exchange of prenyl Rabs 1A, 2, 5 or 9 (not shown), thus it is not a GEF under these conditions.

Gel filtration chromatography revealed that the activity observed reflects the physical displacement of Rab9 from GDI (Fig. 2). Rab–GDI complexes chromatograph at ~80 kDa and can be resolved from free GDI (~55 kDa) and released prenyl Rabs that migrate as a small aggregate at ~100 kDa (ref. 10). In the absence of human Yip3, prenyl Rab9 and GDI co-chromatographed precisely. In contrast, addition of human Yip3 (sequentially as in Fig. 1d) led to the complete dissociation of the complexes analysed.

Liposome-associated Yip3 also catalysed the incorporation of prenyl Rab9 into liposomes from a Rab9–GDI complex (Fig. 3a). The association of prenyl Rab9 with Yip3-containing membranes was specific in that red blood cell-derived proteoliposomes did not serve as acceptors unless they also contained Yip3 protein (Fig. 3a). Moreover, membranes purified from cells overexpressing Yip3 approximately 10-fold were enhanced in their ability to recruit Rab9 from Rab9–GDI complexes (Fig. 3b). Importantly, membrane recruitment reactions containing Yip3-enriched membranes (10 µg) recruited 15-fold more Rab9 (1.5 pmol) than Yip3 present

in those membranes (100 fmol). Finally, anti-Yip3 IgG inhibited the membrane recruitment of Rab9 onto endosome membranes by approximately 40% (Fig. 3c). This provides additional evidence that Yip3 can act catalytically to displace Rab–GDI complexes when incorporated into native endosome membranes.

As Yip3 is localized to the late Golgi¹⁸ and endocytic pathway (Supplementary Information), we tested whether it showed preference in terms of its ability to catalyse GDI displacement from endocytic Rabs compared with endoplasmic reticulum and early Golgi Rabs. Yip3 was active on the early endosome-associated Rab5 and the late endosome-associated Rab9 (Fig. 3d, e) and Rab7 (not shown) but not on Rabs 1A and 2, Rabs of the endoplasmic reticulum and early Golgi (Fig. 3d, e). Although some weak interaction with Rab1A and Rab2 was detected that could reflect direct binding, Yip3 did not act catalytically to displace these Rabs from GDI. This is important, because numerous two hybrid screens have observed broad binding of Yip family members to a variety of Rabs (refs 16, 20). Although Yip3 may bind weakly to multiple Rabs, it appears to act catalytically to displace only a subset of Rabs from GDI. A preference for endocytic Rabs was also seen previously for GDF partially purified from endosomes¹⁰.

The specificity of Yip3 function was confirmed in living cells. RNA interference (RNAi) decreased Yip3 levels ~60% (Fig. 3f). Under these conditions, the percentage of membrane associated Rab9 decreased ~40% while levels of membrane associated Rab1 were unchanged; total cellular Rab1 and Rab9 levels did not change (not shown). This shows that Yip3 acts in some way to influence the membrane association of Rab9 *in vivo*.

These experiments describe the molecular identification of a protein that can act as a Rab–GDI-displacement factor. We favour a model in which an endoplasmic-reticulum-localized Yip protein (Yip1p or PRA-2) drives membrane recruitment of secretory pathway Rabs, whereas the more distally localized Yip3 can recruit endosomal Rabs onto membranes. Both of these classes of proteins may enter the secretory pathway during their biosynthesis; Yip1p and PRA-2 may be retained in the endoplasmic reticulum while Yip3 is transported to the plasma membrane and thereby, the

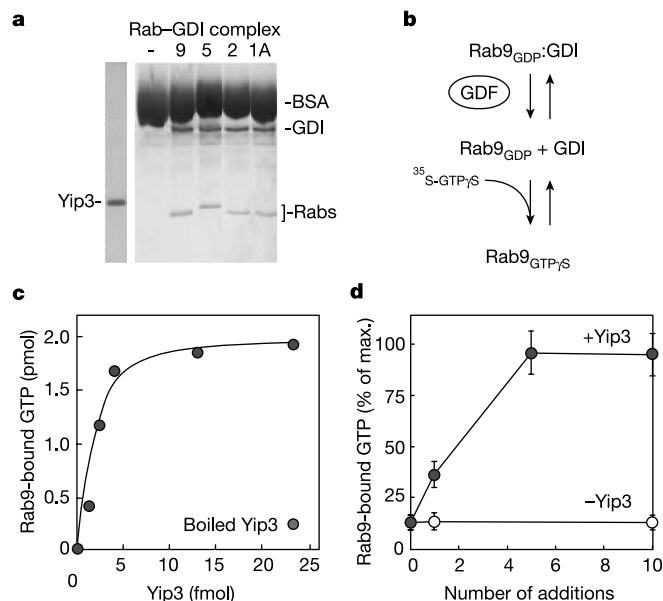


Figure 1 Yip3 dissociates Rab9–GDI complexes catalytically. **a**, Coomassie stained SDS–PAGE of His-tagged human Yip3 (1 µg) and untagged Rab–GDI complexes (300 ng Rab) on a layered gel (6%/12.5%). Complexes were stabilized by BSA (30 µg). **b**, Nucleotide binding assay for GDI displacement from Rab–GDI complexes. **c**, Yip3 dissociates Rab9–GDI complexes as measured by Rab9 nucleotide exchange. Complexes were incubated with Yip3 for 20 min at 37 °C. **d**, Yip3 (47 fmol) or buffer was added once, every 4 min or every 2 min at 37 °C. Data represent averages of triplicates (error ~15%). Maximum nucleotide exchange (**c**, 7.9 pmol; **d**, 3.6 pmol) was determined by disruption of complexes in 1% CHAPS.

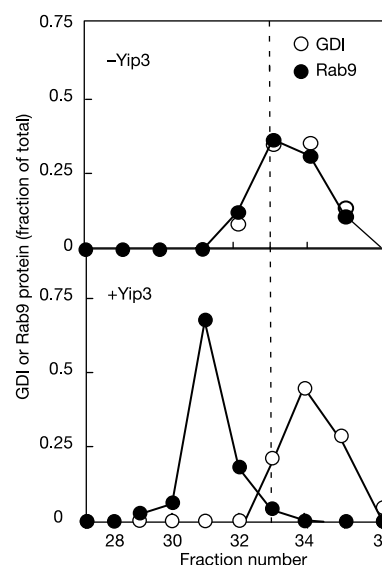


Figure 2 Human Yip3 catalyses the physical displacement of Rab9 from GDI, as determined by gel filtration chromatography. Prenyl Rab9–GDI complexes (15.6 pmol) were incubated without (top panel) or with pure Yip3 (0.94 pmol total, bottom panel) and analysed by Superose 12 gel filtration chromatography. Rab9 (filled circles) and GDI (open circles) were determined by immunoblotting. To ensure efficient reaction completion, Yip3 (or buffer) was added sequentially in 47-fmol aliquots as in Fig. 1d.

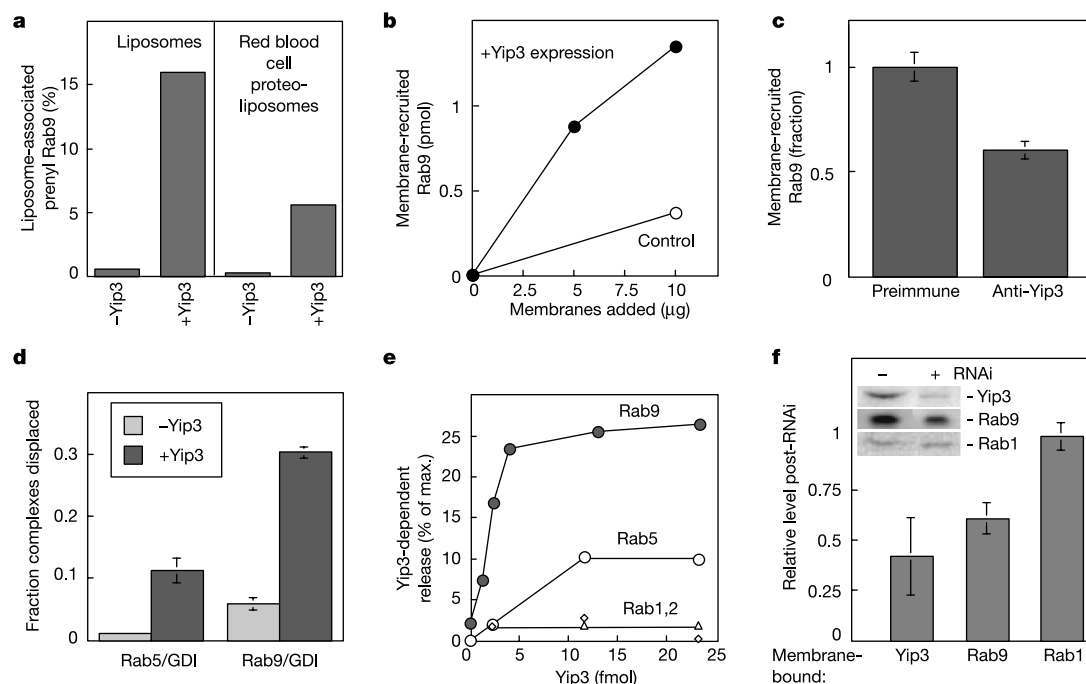


Figure 3 Yip3-catalysed membrane recruitment is specific for endosomal Rabs *in vitro* and *in vivo*. **a**, Rab9 recruitment from Rab9–GDI complexes into liposomes (left) or red blood cell ghost-proteoliposomes (right) with (+Yip3) or without (–Yip3) recombinant human Yip3. **b**, Rab9 recruitment from Rab9–GDI complexes (4.1 pmol) onto membranes from HEK293 cells expressing (filled circles) or not expressing (open circles) human Yip3. Membranes contained identical amounts of late endosomes as monitored by 300 kDa mannose 6-phosphate receptor immunoblot. **c**, Anti-Yip3 IgG (0.83 $\mu\text{g ml}^{-1}$) but not preimmune IgG inhibits recruitment of prenyl Rab9 onto rat liver endosomes (12 $\mu\text{g ml}^{-1}$). **d**, Human Yip3 (0.47 pmol) specifically dissociates endosomal Rab–GDI

complexes (~ 9 pmol Rab) (see Fig. 1b). **e**, Yip3 dissociates endosomal, Rab5– and Rab9–GDI complexes but not endoplasmic reticulum and Golgi-associated, Rab1A– and Rab2–GDI complexes. Incubations were as in **d**. Maximum exchange was 6.8, 6.1, 7.6 and 7.9 pmol GTP γ S for Rab1A, Rab2, Rab5 and Rab9; values were normalized to 100%. Data represent triplicate measurements (error $\sim 15\%$). **f**, Yip3 RNAi decreases membrane-associated Rab9 but not Rab1. Inset, immunoblot. Controls contained luciferase RNAi; protein levels were normalized for membrane recovery using transferrin receptor, determined by immunoblot.

endocytic pathway. Rabs could use these Yips to associate with the appropriate pathway, and then become stabilized in a given compartment by interaction with specific effector proteins.

Yeast Yip1p occurs as a stoichiometric complex with Yif1p^{14,21}. It will be important to determine the oligomeric nature of Yip3 in living cells, and also to determine if Yip heterodimers show enhanced catalytic activity or specificity. Alternatively, different Yip family member combinations may display differences in localization to help achieve the distinct localizations of Rab GTPases in eukaryotic cells. Depletion of Yip1p and Yif1p did not decrease membrane association of yeast Rab1 homologue, Ypt1p²². Another Yip may have catalysed Ypt1p recruitment under these conditions. Yips interact with SNARE proteins and coat proteins^{15,22,23}. Much work remains to determine their broad roles in eukaryotic cells. □

Methods

Proteins and Rab–GDI complexes

Human Yip3 was amplified from a human testis complementary DNA preparation by polymerase chain reaction (PCR) and cloned. Amino-terminal His-tagged Yip3 was expressed in *E. coli* BL-21 (DE3) RIL cells ($\text{OD}_{600} = 0.4$) induced with 1 mM IPTG for 2 h, 30 °C. Cells (500-ml culture) were resuspended in 30 ml of 25 mM HEPES pH 7.5, 300 mM KCl, 1 mM EDTA, 1 mM DTT, 1 mM PMSF, 1 μM pepstatin, 2.1 mM leupeptin and 0.14 TIU ml^{-1} aprotinin, and broken by French pressure cell. (Some soluble Yip3 is produced in *E. coli* and mammalian cells—it is an aggregate $>600,000$ K_D , and was discarded.) Membrane pellets were solubilized in 25 ml of 5% Triton X-100, 25 mM Tris/HCl, pH 8.0, 1 mM EDTA, 1 mM DTT and 1 mM PMSF in a Potter–Elvehjem homogenizer with mixing for 15 min, spun 45 min at 100,000g (60Ti rotor, Beckman). The supernatant was applied onto a Q-Sepharose (Pharmacia) column (2 ml) equilibrated in 25 mM Tris/HCl, 10% glycerol and 1% Triton X-100. Yip3 was eluted with a 0–500 mM NaCl gradient (40 ml) and applied onto a Ni-NTA agarose (Qiagen) column (1 ml) equilibrated in 25 mM HEPES/NaOH pH 7.5, 300 mM KCl, 50 mM imidazole and 1% CHAPS. Bound Yip3 was washed with 20 ml buffer to remove Triton X-100, then eluted

with a 50–400 mM imidazole gradient (20 ml). Yip3 fractions were pooled and concentrated in a 10 kDa Centricon concentrator (Amicon) and dialysed against 25 mM HEPES, pH 7.5, 100 mM NaCl (buffer Y) + 1% CHAPS. Rabbit anti-Yip3 antiserum was prepared using His-tagged Yip3 in Freund's adjuvant; IgG was purified by protein A-agarose. Serum was affinity purified using GST–Yip3(1–79).

Prenylated Rabs were purified from insect cell membranes²⁴. Bovine GDI was from frozen brain⁴. Prenylated Rab–GDI complexes²⁴ were stored at -80 °C with 1 mg ml^{-1} BSA (Boehringer). Nucleotide exchange and GDI displacement were as described¹⁰. Rab9–GDI complexes were re-purified after thawing by centrifugation for 10 min at 321,000g in a TLA 120.1 rotor; supernates (250 μl) were applied onto a Superose 12 column (24 ml, Pharmacia) equilibrated in 64 mM HEPES/NaOH pH 8.0, 100 mM NaCl, 8 mM MgCl_2 , 2 mM EDTA, 10 μM GDP and 0.2 mM DTT. Fractions (0.4 ml) were used in subsequent nucleotide exchange assays. The reaction mixture was loaded directly onto a Superose 12 column equilibrated and run in the above buffer with 10 μM GTP γ S instead of GDP.

Liposome and membrane recruitment assays

His-tagged Yip3 was incorporated into liposomes (soybean lipids, Avanti Polar Lipids) by mixing 100 μl Yip3 (0.1 mg ml^{-1}) and 100 μl lipid (2 mM) in buffer Y + 1% CHAPS, followed by removal of CHAPS by dialysis. Lipid yield was measured by phospholipid assay²⁵; Yip3 yield was 60% by SDS–polyacrylamide gel electrophoresis (SDS–PAGE). Anti-His tag immunoblot analysis of Nycodenz flotation gradient fractions showed Yip3 to be quantitatively liposome-incorporated. For red blood cell ghost proteoliposomes²⁶, Yip3 (50 $\mu\text{g ml}^{-1}$), ghosts (0.5 mg ml^{-1} , ~ 0.5 mM phospholipid) and 1.0 mg ml^{-1} soybean lipid were mixed in 4% CHAPS and dialysed as above.

Rab9 recruitment onto liposomes was measured by incubating liposomes (7.5 nM Yip3, 50 μM lipid) and Rab9–GDI (136 nM), or Yip3-containing proteoliposomes (4.4 nM Yip3, 50 μM lipid) and Rab9–GDI (42 nM), in buffer Y plus 100 mM $(\text{NH}_4)_2\text{SO}_4$, 1 mM DTT, 30 μM GTP γ S and 6.2 mM $\text{Mg}(\text{OAc})_2$ at 37 °C for 20 min. Liposome- and GDI-associated Rab9 were resolved on a Nycodenz flotation gradient: each 100 μl reaction was mixed with 300 μl ice-cold 48% Nycodenz (w/v) in buffer A (buffer Y + 30 μM GTP γ S, 1.2 mM $\text{Mg}(\text{OAc})_2$), and overlaid with 250 μl , 30% Nycodenz + buffer A, followed by 100 μl of buffer A. Gradients were spun at 342,000g in a TLA-100.2 rotor (Beckman) for 1 h at 4 °C. Soluble proteins remained in the bottom 450 μl ; liposomes and associated proteins were in the top 300 μl . Fractions (150 μl) were analysed by anti-Rab9 and anti-GDI α immunoblot.

Mammalian cell overexpression and siRNA depletion

Yip3 was cloned into the mammalian vector pcDNA3.1. HEK293 cells were transfected with pcDNA3.1 human Yip3 or pcDNA3.1, collected 42 h later, lysed by osmotic shock and shearing, and membranes were purified by ultracentrifugation. Membrane recruitment was as published^{6,7} in buffer plus 100 mM (NH₄)₂SO₄, 100 μM GTPγS, 0.1 mg ml⁻¹ BSA, protease inhibitor cocktail (0.34 units ml⁻¹ aprotinin, 0.01 mg ml⁻¹ leupeptin, 1 μM pepstatin A), 0.2 mM DTT and ATP regeneration system^{6,7}. Rab9-GDI complex (300 ng, 4.1 pmol) and membranes were mixed in 250 μl for 40 min at 37 °C. Ice-cold buffer (64 mM HEPES-KOH, pH 8.0, 100 mM NaCl, 8 mM MgCl₂, 2 mM EDTA; 0.5 ml) was added and each reaction was then layered over 0.4 M sucrose + ice-cold buffer and spun for 10 min at 342,000g, 2 °C in a TLA-100.2 rotor (Beckman). Pellets were analysed by 12.5% SDS-PAGE and anti-Rab9 immunoblot.

HeLa S-3 cells were transfected with 0.13 μM siRNA duplex (sense 5'-(GCUUGUG CUCUUUGGCCGA)d(TT)-3'; Dharmacon) using Oligofectamine (Invitrogen). After 77 h, cells were lysed in 25 mM HEPES pH 7.3 with protease inhibitors by 10 passages through a 25G needle. Extracts were spun 30s at 1,000 g; supernatants were centrifuged 15 min, 321,000g in a TLA 120.1 rotor (Beckman). The membrane pellet was washed in PBS.

Received 16 July; accepted 8 September 2003; doi:10.1038/nature02057.

1. Zerial, M. & McBride, H. Rab proteins as membrane organizers. *Nature Rev. Mol. Cell Biol.* **2**, 107–117 (2001).
2. Segev, N. Ypt and Rab GTPases: Insight into functions through novel interactions. *Curr. Opin. Cell Biol.* **13**, 500–511 (2001).
3. Pfeffer, S. R. Rab GTPases: Specifying and deciphering organelle identity and function. *Trends Cell Biol.* **11**, 487–491 (2001).
4. Sasaki, T. *et al.* Purification and characterization from bovine brain cytosol of a protein that inhibits the dissociation of GDP from and the subsequent binding of GTP to smg p25A, a ras p21-like GTP-binding protein. *J. Biol. Chem.* **265**, 2333–2337 (1990).
5. Pfeffer, S. R., Dirac-Svestrup, A. B. & Soldati, T. Rab GDP dissociation inhibitor: Putting Rab GTPases in the right place. *J. Biol. Chem.* **270**, 17057–17059 (1995).
6. Soldati, T., Shapiro, A. D., Svestrup, A. B. & Pfeffer, S. R. Membrane targeting of the small GTPase Rab9 is accompanied by nucleotide exchange. *Nature* **369**, 76–78 (1994).
7. Soldati, T., Rancano, C., Geissler, H. & Pfeffer, S. R. Rab7 and Rab9 are recruited onto late endosomes by biochemically distinguishable processes. *J. Biol. Chem.* **270**, 25541–25548 (1995).
8. Ullrich, O., Horiuchi, H., Bucci, C. & Zerial, M. Membrane association of Rab5 mediated by GDP-dissociation inhibitor and accompanied by GDP/GTP exchange. *Nature* **368**, 157–160 (1994).
9. Shapiro, A. D. & Pfeffer, S. R. Quantitative analysis of the interactions between prenyl Rab9, GDP dissociation inhibitor-α, and guanine nucleotides. *J. Biol. Chem.* **270**, 11085–11090 (1995).
10. Dirac-Svestrup, A. B., Sumizawa, T. & Pfeffer, S. R. Identification of a GDI displacement factor that releases endosomal Rab GTPases from Rab-GDI. *EMBO J.* **16**, 465–472 (1997).
11. Yang, X., Matern, H. T. & Gallwitz, D. Specific binding to a novel and essential Golgi membrane protein (Yip1p) functionally links the transport GTPases Ypt1p and Ypt31p. *EMBO J.* **17**, 4954–4963 (1998).
12. Calero, M. & Collins, R. N. S. *cerevisiae* Pra1p/Yip3 interacts with Yip1p and Rab proteins. *Biochem. Biophys. Res. Commun.* **290**, 676–681 (2002).
13. Calero, M., Winand, N. J. & Collins, R. N. Identification of the novel proteins Yip4p and Yip5p as Rab GTPase interacting factors. *FEBS Lett.* **515**, 89–98 (2002).
14. Matern, H. *et al.* A novel Golgi membrane protein is part of a GTPase-binding protein complex involved in vesicle targeting. *EMBO J.* **19**, 4485–4492 (2000).
15. Martincic, I., Peralta, M. E. & Ngsee, J. K. Isolation and characterization of a dual prenylated Rab and VAMP2 receptor. *J. Biol. Chem.* **272**, 26991–26998 (1997).
16. Bucci, C., Chiariello, M., Lattero, D., Maiorano, M. & Bruni, C. B. Interaction cloning and characterization of the cDNA encoding the human prenylated Rab acceptor (PRA1). *Biochem. Biophys. Res. Commun.* **258**, 657–662 (1999).
17. Hutt, D. M., Da-Silva, L. F., Chang, L. H., Prosser, D. C. & Ngsee, J. K. PRA1 inhibits the extraction of membrane-bound Rab GTPase by GDI1. *J. Biol. Chem.* **275**, 18511–18519 (2000).
18. Abdul-Ghani, M., Gougeon, P. Y., Prosser, D. C., Da-Silva, L. F. & Ngsee, J. K. PRA isoforms are targeted to distinct membrane compartments. *J. Biol. Chem.* **276**, 6225–6233 (2001).
19. Lin, J., Liang, Z., Zhang, Z. & Li, G. Membrane topography and topogenesis of prenylated Rab acceptor (PRA1). *J. Biol. Chem.* **276**, 41733–41741 (2001).
20. Figueroa, C., Taylor, J. & Vojtek, A. B. Prenylated Rab acceptor protein is a receptor for prenylated small GTPases. *J. Biol. Chem.* **276**, 28219–28225 (2001).
21. Otte, S. *et al.* Erv41p and Erv46p: New components of COPII vesicles involved in transport between the ER and Golgi complex. *J. Cell Biol.* **152**, 503–518 (2001).
22. Barrowman, J., Wang, W., Zhang, Y. & Ferro-Novick, S. The Yip1p-Yif1p complex is required for the fusion competence of endoplasmic reticulum-derived vesicles. *J. Biol. Chem.* **278**, 19878–19884 (2003).
23. Tang, B. L. *et al.* A membrane protein enriched in ER exit sites interacts with COPII. *J. Biol. Chem.* **276**, 40008–40017 (2001).
24. Soldati, T., Shapiro, A. D. & Pfeffer, S. R. Reconstitution of the endosomal targeting of rab9 protein using purified, prenylated rab9 protein as a complex with GDI. *Methods Enzymol.* **257**, 253–259 (1995).
25. Lanzetta, P. A., Alvarez, L. J., Reinach, P. S. & Candia, O. A. An improved assay for nanomole amounts of inorganic phosphate. *Anal. Biochem.* **100**, 95–97 (1979).
26. Fairbanks, G., Steck, T. L. & Wallach, D. F. H. Electrophoretic analysis of the major polypeptides of the human erythrocyte membrane. *Biochemistry* **22**, 2606–2617 (1971).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank F. Barr and T. Meyer for providing Rab2 and Rab1A plasmids, respectively, and B. Stafford and C. Melton for technical assistance. This work was supported by grants from the NIH and the American Heart Association. U.S. and D.A. were postdoctoral fellows of the Swedish Foundation for International Cooperation in Research and Higher Education and the Leukemia and Lymphoma Society, respectively.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.R.P. (pfeffer@stanford.edu).

Targets of the cyclin-dependent kinase Cdk1

Jeffrey A. Ubersax¹, Erika L. Woodbury¹, Phuong N. Quang¹, Maria Paraz¹, Justin D. Blethrow^{1,2}, Kavita Shah³, Kevan M. Shokat² & David O. Morgan¹

¹Departments of Physiology and Biochemistry & Biophysics, and ²Department of Cellular and Molecular Pharmacology, University of California, San Francisco, California 94143, USA

³Genomics Institute of the Novartis Foundation, San Diego, California 92121, USA

The events of cell reproduction are governed by oscillations in the activities of cyclin-dependent kinases (Cdks)¹. Cdks control the cell cycle by catalysing the transfer of phosphate from ATP to specific protein substrates. Despite their importance in cell-cycle control, few Cdk substrates have been identified². Here, we screened a budding yeast proteomic library for proteins that are directly phosphorylated by Cdk1 in whole-cell extracts. We identified about 200 Cdk1 substrates, several of which are phosphorylated *in vivo* in a Cdk1-dependent manner. The identities of these substrates reveal that Cdk1 employs a global regulatory strategy involving phosphorylation of other regulatory molecules as well as phosphorylation of the molecular machines that drive cell-cycle events. Detailed analysis of these substrates is likely to yield important insights into cell-cycle regulation.

We developed methods for the systematic identification of Cdk substrates in the budding yeast *Saccharomyces cerevisiae*, a simple eukaryote in which the cell cycle is controlled by a single Cdk, Cdk1 (or Cdc28). We used a recently developed approach that allows the specific labelling of the substrates of a single kinase in a cell extract^{3,4}. This method involves the mutation of a conserved bulky residue in the ATP-binding pocket to a glycine or alanine. The resulting analogue-sensitive (as) kinase is able to bind bulky ATP analogues that fit into the mutant-binding pocket but cannot bind to wild-type kinases. Importantly, the mutation is deep in the ATP-binding pocket, far from the protein substrate binding site, and is not expected to alter the protein substrate specificity of the protein kinase⁵. The addition of radiolabelled ATP analogue to a

Table 1 Nucleotide specificity of Cdk1-as1

Kinase	Nucleotide	K_m (μM)	k_{cat} (min ⁻¹)	k_{cat}/K_m (μM ⁻¹ min ⁻¹)
Cdk1-Clb2	ATP	35	132	3.73
Cdk1-as1-Clb2	ATP	322	21.3	0.07
Cdk1-Clb2	N ⁶ -(benzyl) ATP	>1,000	NM	NM
Cdk1-as1-Clb2	N ⁶ -(benzyl) ATP	1.5	13.6	9.07

Active cyclin-Cdk complexes were formed by the addition of excess MBP-Clb2 to purified Cdk1-His₆ or Cdk1-as1-His₆, and their ability to phosphorylate histone H1 at different N⁶-(benzyl) ATP or ATP concentrations was measured to generate K_m and k_{cat} values. NM, V_{max} not measurable at 1 mM nucleotide.

cell extract containing a single as-kinase therefore leads to the specific labelling of the direct substrates of that kinase.

In previous work, we constructed an analogue-sensitive version of yeast Cdk1, Cdk1-as1, by replacing phenylalanine 88 with glycine⁶. We showed that Cdk1-as1 is functional *in vivo* and is uniquely sensitive to a bulky chemical inhibitor that interacts with the enlarged ATP-binding site. In the present work, we measured Cdk1-as1 activity with the bulky ATP analogue N⁶-(benzyl) ATP. As predicted, Cdk1-as1 (in a complex with the mitotic cyclin Clb2) exhibited a high affinity for the ATP analogue, whereas wild-type Cdk1 had no measurable activity with this substrate (Table 1). Cdk1-as1 was approximately 130-fold more active towards N⁶-(benzyl) ATP than ATP, that is, $(k_{cat}/K_m)^{N^6-(benzyl)ATP} / (k_{cat}/K_m)^{ATP} = 130$ (Table 1). Thus, the substitution of a glycine for a phenylalanine in the active site of Cdk1 results in a mutant that displays high affinity and selectivity for N⁶-(benzyl) ATP.

Concentrated cell extracts, prepared by gentle detergent lysis of yeast spheroplasts, were incubated with radiolabelled N⁶-(benzyl) ATP. Few proteins were labelled, indicating that wild-type kinases in the extract were not using the ATP analogue at a significant rate (Fig. 1, lane 1). We then added purified Cdk1-as1-Clb2 complexes at a concentration (7 nM) that resulted in levels of activity similar to those in mitotic cell extracts. Many radiolabelled proteins were generated (Fig. 1, lane 2). Incubation of purified Cdk1-as1-Clb2 with ATP analogue in the absence of cell extract resulted in autophosphorylation of Clb2 and other contaminants (Fig. 1, lane 3), and revealed that most of the proteins labelled in lane 2 represent direct Cdk1 substrates in the crude cell extract.

We next developed a simple, rapid and highly sensitive approach to scan the yeast proteome for proteins phosphorylated by Cdk1-as1-Clb2. We used a library of 6,144 yeast strains, each expressing a unique open reading frame (ORF) with an amino-terminal fusion

to glutathione-S-transferase (GST)⁷. We reasoned that we could perform kinase reactions with Cdk1-as1 in lysates of strains from this library, and then purify the GST-tagged protein to determine if it was phosphorylated.

Pilot studies revealed that tagged substrate phosphorylation was not readily detected if reactions were performed with extracts made from a pool of multiple GST-ORF yeast strains. It was therefore necessary to perform separate reactions for each GST-tagged protein. Rather than performing over six thousand reactions, we focused on several hundred candidate ORFs that seemed likely to encode Cdk targets. As most known Cdk substrates contain multiple Cdk consensus phosphorylation sites (S/T*-P-x-K/R, where x = any amino acid), we first tested the 385 ORFs encoding proteins with two or more of these sites. In addition, because cell-cycle regulators are often controlled by transcriptional mechanisms, we screened the 137 ORFs encoding proteins with a single Cdk consensus site and whose transcripts are cell-cycle regulated⁸. Finally, in an attempt to estimate the number of Cdk targets in the entire proteome, we tested 198 randomly chosen ORFs. These candidates encompass a total of 695 unique ORFs (due to 25 overlaps between the random 198 and the 522 candidates), which represent about 11% of the yeast proteome.

Results from twelve representative reactions are shown in Fig. 2a. To rank the substrates, we measured the phosphorylation of each tagged protein (and fragments) and divided by the amount of tagged protein present in the reaction. The amounts of phosphate incorporated per nanogram of protein ranged over seven orders of magnitude. The logarithms of these values were calculated and defined as the 'P-score' for each tagged protein.

Of the 695 tested proteins, 360 were detectably phosphorylated (Fig. 2b). Of these 360 proteins, 181 had P-scores of 2.0 or higher and represent the most interesting group of substrates because they were less-abundant proteins with high levels of phosphorylation. As shown in Fig. 2c, 179 of the 181 top substrates were present in our reactions at very low concentrations (<50 pM to 50 nM; see Fig. 2c legend), suggesting that phosphorylation of these proteins was not an artefact of high substrate concentrations. Forty of the top 181 substrates are listed in Table 2 (a complete list of results with all tested proteins is available as Supplementary Information).

Approximately 12 proteins containing full Cdk consensus sites have been identified previously as likely Cdk targets in budding yeast (Table 2, marked in bold: Swi5, Sic1, Cln2, Cdh1, Far1, Gin4, Swi1, Cdc6, Orc2, Orc6, Sld2 and Pds1)^{9–20}. Ten of these proteins had P-scores greater than 2 in our screen. Phosphorylation of the remaining two, Cdc6 and Cln2, was apparently undetectable as a result of the limitations of the tagged protein library (see Methods). Thus, essentially all known Cdk targets with full consensus sites are among the top 181 substrates we identified, which strongly supports the validity of our approach.

Our results provide insights into mechanisms that govern substrate recognition by Cdks. A small number of Cdk targets in yeast and other species do not contain the full Cdk consensus site (S/T*-P-x-K/R) but are instead phosphorylated at a minimal consensus site containing a serine or threonine residue followed by proline (S/T*-P)²¹. Of the 198 random proteins we tested, 16 were phosphorylated with P-scores of greater than two. Thirteen of these proteins contain the full consensus sequence, and the remaining three contain multiple copies of the minimal consensus site. We therefore suspect that the great majority of Cdk targets contain the full consensus site. However, the presence of a full consensus site is not sufficient for Cdk phosphorylation: we found 123 proteins with this site that were not phosphorylated, despite being expressed at detectable levels.

Substrate recognition by Cdks is influenced by the associated cyclin subunit^{22,23}. Although our screen was carried out with the mitotic Cdk1-as1-Clb2 complex, we found that several substrates involved in DNA replication (including Orc6, Mcm3 and Sld2) are

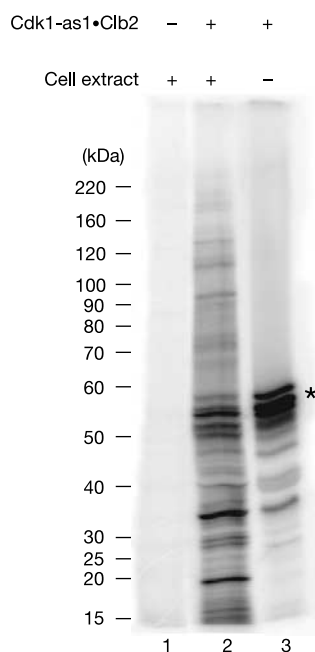


Figure 1 Addition of Cdk1-as1 and radiolabelled N⁶-(benzyl) ATP to a cell extract results in the appearance of many specific phosphoproteins. 5 μ Ci γ -³²P-N⁶-(benzyl) ATP was added to 30 μ g whole yeast extract (lanes 1, 2) or buffer (lane 3) in the presence (lanes 2, 3) or absence (lane 1) of 7 nM purified Cdk1-as1-Clb2. After incubation for 30 min at room temperature (23 °C), reactions were analysed by SDS-PAGE on a 5–15% gradient gel, and autoradiographed. Asterisk indicates autophosphorylated Clb2 bands.

phosphorylated tenfold more rapidly by the S-phase Cdk1-Clb5 complex (M. Loog and D.O.M., unpublished results). We therefore believe that our screen has led to the identification of a broad range of Cdk targets involved in multiple cell-cycle stages.

To confirm that we have identified proteins that are substrates of Cdk1 in the cell, it is important to show that phosphorylation of these proteins *in vivo* is dependent on Cdk1 activity. Traditionally, temperature-sensitive *cdk1* mutants have been used to demonstrate that phosphorylation is Cdk1-dependent *in vivo*. However, inactivation of these mutants takes 2 to 3 h at the restrictive temperature and results in a cell-cycle block (typically in G1), making it difficult to rule out indirect effects caused by the shift in cell-cycle position. These problems can be circumvented with the *cdk1-as1* strain, in which it is possible to inhibit much of the cell's Cdk1 activity within 5–10 min by addition of high concentrations (5–25 μ M) of 1-NM-PP1, a small molecule that inhibits the analogue-sensitive mutant Cdk1 without affecting other kinases in the cell⁶. The speed of inhibition reduces the likelihood of indirect effects.

To rapidly assess the phosphorylation state of a subset of substrates *in vivo*, we identified candidates that migrate on poly-

acrylamide gels as multiple bands (such mobility shifts can be a useful indication of phosphorylation). We prepared or obtained strains expressing epitope-tagged versions of 171 of the 181 candidate substrates expressed at their endogenous loci. A hundred proteins were detectable on western blots, and approximately 35 of these proteins displayed reproducible mobility shifts. We initially focused on Slk19, a non-essential regulator of spindle assembly and stability. Slk19 is a substrate of separase (Esp1)²⁴ and migrates on gels as a pair of band clusters, in which the lower cluster is the carboxy-terminal separase-cleavage product. As reported previously²⁴, the reduced mobility of the upper forms in each cluster is caused by phosphorylation and is maximal in cells arrested in S phase or mitosis (Fig. 3a), consistent with phosphorylation by Cdk1. Treatment of *cdk1-as1* cells with 1-NM-PP1 caused dephosphorylation of the upper Slk19 band within five minutes, arguing that phosphorylation *in vivo* is Cdk1-dependent (Fig. 3b). After 30 min of treatment, the drop in mitotic cyclin levels presumably reflects activation of the Cdh1-dependent APC, which is normally inhibited by Cdk1^{12,13}. Phosphorylation of the separase-cleavage product was not reversed by Cdk1 inhibition, perhaps indicating

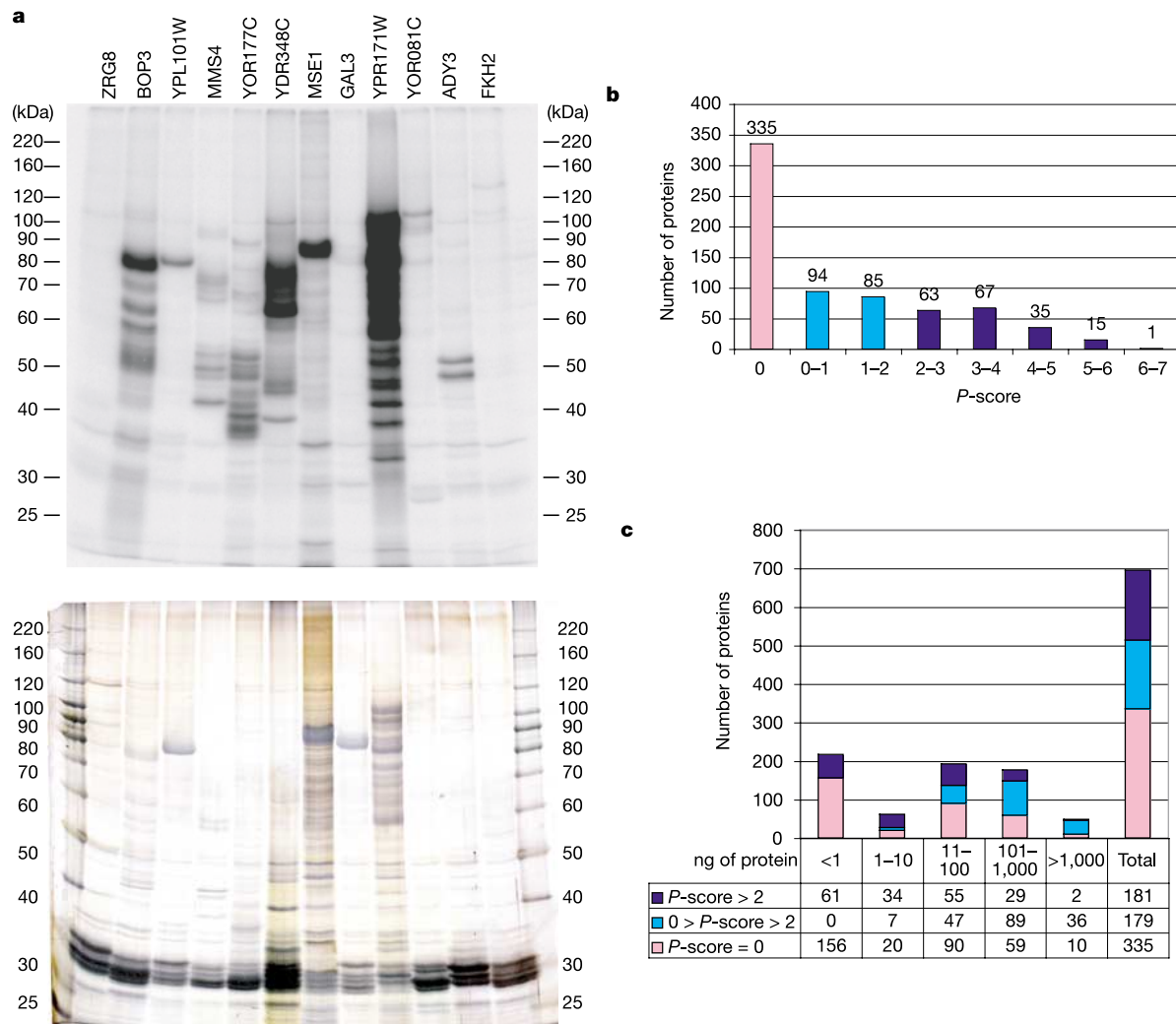


Figure 2 181 proteins are efficiently phosphorylated by Cdk1-as1-Clb2. **a**, Autoradiograph (top) and silver-stained gel (bottom) of reactions with twelve GST-fusion proteins. 7 nM Cdk1-as1-Clb2 and 5 μ Ci γ -³²P-N⁶-(benzyl) ATP were added to whole yeast extracts expressing a single GST-tagged protein and incubated for 30 min at room temperature, after which the GST-tagged protein was purified and analysed by SDS-PAGE. **b**, Histogram of the distribution of *P*-scores for the 695 GST-tagged proteins

reveals 181 efficiently phosphorylated Cdk1 substrates (*P*-score > 2, dark blue). **c**, Most of the best Cdk1 substrates are present at low concentrations in our reactions. GST-tagged proteins were grouped by their estimated protein amounts (nanograms in a 200 μ l reaction volume). If we assume a relative molecular mass of 100,000, then 1,000 ng tagged protein represents a concentration of 50 nM. The number of proteins with *P*-scores above 2, between 0 and 2, or equal to zero, are shown for each group.

that this fragment is less susceptible to phosphatases or is phosphorylated by other kinases.

Slk19 contains three full Cdk consensus sites (S/T*-P-x-K/R) and three minimal Cdk consensus sites (S/T*-P). Changing all six of these residues to alanine abolished phosphorylation of the protein *in vivo* (data not shown), further arguing that Slk19 is phosphorylated by Cdk1 *in vivo*. Replacement of the endogenous *SLK19* gene with a mutant gene encoding the non-phosphorylatable mutant resulted in a strain that appeared wild-type in all respects (cell-cycle timing, spindle structure and sporulation), indicating that Slk19 phosphorylation at these sites is not essential for cell-cycle progression under routine laboratory conditions.

We next analysed the effects of Cdk1 inhibition on other candidates with detectable gel mobility shifts. A fifteen-minute treatment of asynchronous *cdk1-as1* cells with 1-NM-PP1 caused a complete or partial collapse in the mobility of 11 novel Cdk1 substrates in addition to Slk19, arguing that phosphorylation of these proteins *in vivo* is Cdk1-dependent (Fig. 3c).

We also analysed two previously identified Cdk1 substrates, Orc6¹⁸ and Pds1²⁰. These proteins, as well as some of our candidates, were only partially dephosphorylated upon inhibitor addition (Fig. 3c). An underlying assumption in all approaches involving kinase inhibition *in vivo* is that phosphates on proteins are rapidly turned over in the cell. The steady-state phosphate content at any

phosphorylation site is determined by the relative rates of the protein kinase(s) and protein phosphatase(s) acting on that site. For some substrates, it is conceivable that kinase activity is present in 100- or 1,000-fold excess over phosphatase activity, in which case 90 or 99% inhibition of the kinase might not lead to a detectable drop in the phosphate content of the substrate. Thus, this method will work only when the relevant phosphatase activity is present in significant amounts.

Our results suggest that a significant fraction of the 181 candidate substrates are Cdk1 targets *in vivo*. Because of limitations in our analysis, however, it is not yet possible to accurately estimate the total number of bona fide Cdk1 targets on our list. New methods will be required to allow more rapid analysis of the phosphorylation state *in vivo* of large numbers of proteins in the presence and absence of Cdk1 activity. Recently, for example, Ficarro *et al.*²⁵ used mass spectrometry to sequence large numbers of phosphopeptides from yeast extracts. Peptides from three of our 181 best substrates (Mcm3, Mob1 and YDL113C) were found to be phosphorylated at Cdk consensus sites in their analysis. Similar studies with lysates of *cdk1-as1* cells, with and without 1-NM-PP1, could provide a global approach to validating Cdk1 substrate candidates.

About one-third of the 181 best substrates, like those listed in Table 2, are involved in processes that are regulated during the cell cycle. Many of these processes are governed by Cdk1, but the

Table 2 **Forty selected Cdk1 substrates**

Protein*	Peak†	Full Cdk consensus sites	Estimated protein level (ng)‡	Phosphorylation (arbitrary units)	P-score	Biological process
Cdk1 regulation						
Cdh1		6	25	304,756	4.1	APC activating factor in M/G1
Cln2	G1	1	30	0	0.0	G1-specific cyclin controlling events at START
Far1	G2/M	4	100	65,166	2.8	Mating-specific Cdk1 inhibitor
Mih1		2	5	35,512	3.9	Phosphatase that dephosphorylates Y19 on Cdk1
Sic1	M/G1	3	200	996,630	3.7	Clb-specific Cdk1 inhibitor
Swe1	G1	2	2	87,408	4.6	Protein kinase that phosphorylates Y19 on Cdk1
Swi5	G2/M	8	0.1	5,206	4.7	G1-specific transcription factor
DNA replication						
Cdc6	M/G1	5	0.1	0	0.0	Required for pre-replicative complex formation
Dbf4		2	0.1	5,310	4.7	Regulatory subunit of the protein kinase Cdc7
Eco1	G1	1	30	127,694	3.6	Required for establishment of sister chromatid cohesion
Mcm3	M/G1	5	300	372,637	3.1	Required for initiation of DNA replication
Orc2		6	200	730,607	3.6	Required for pre-replicative complex formation
Orc6		4	50	588,875	4.1	Required for pre-replicative complex formation
Sld2	G1	5	50	51,144	3.0	Required for initiation of DNA replication
Smc4		5	0.1	1,373	4.1	Subunit of the condensin protein complex
Mitosis						
Cdc20	G2/M	1	2.5	6,138	3.4	APC-activating factor in anaphase
Cdc5	G2/M	1	5	7,404	3.2	Polo-like kinase, promotes anaphase and mitotic exit
Dbf2	G2/M	2	50	80,255	3.2	Protein kinase required for mitotic exit
Dbf20	S/G2	1	20	31,795	3.2	Protein kinase required for mitotic exit
Fkh2		3	0.1	2,735	4.4	Transcription factor regulating G2/M gene expression
Lte1		8	0.1	3,749	4.6	GDP exchange factor involved in mitotic exit
Mob1	G2/M	2	50	305,565	3.8	Regulatory subunit of Dbf2, required for mitotic exit
Ndd1	G1	4	10	24,327	3.4	Interacts with Fkh2 to control G2/M gene expression
Net1		3	0.1	4,452	4.6	Regulator of the phosphatase Cdc14
Pds1	G1	3	25	32,213	3.1	Regulator of Esp1 (separase)
Spindle assembly						
Ase1	G2/M	7	0.1	555	3.7	Microtubule binding protein required for spindle stability
Kar3	G1	2	75	182,904	3.4	Kinesin-like protein required for proper spindle assembly
Kip2	S/G2	2	150	91,993	2.8	Kinesin-like protein required for proper spindle assembly
Kip3	S/G2	1	50	7,592	2.2	Kinesin-like protein required for proper spindle assembly
Slk19	G1	3	0.1	17,857	5.3	Kinetochores protein required for spindle assembly
Stu2	S	1	50	5,178	2.0	SPB component regulating microtubule dynamics
Actin polarization						
Bem1	G2/M	2	100	44,466	2.6	Protein required for proper cell polarization
Bem3		5	0.1	596	3.8	GTPase-activating protein for Rho1 & Cdc42
Bni1		3	0.1	15,168	5.2	Protein involved in Rho protein signal transduction
Other process						
Ash1	M/G1	7	5	2,013	2.6	Daughter-cell-localized transcription factor
Bbp1	G1	1	10	13,328	3.1	Spindle-pole-body-associated protein
Chs2	G2/M	4	50	2,744,790	4.7	Chitin synthase II required for proper septum formation
Gin4	G1	2	0.1	781	3.9	Protein kinase required for septin organization
Rad9		9	0.1	604	3.8	DNA damage checkpoint protein
Rts1		2	0.1	2,281	4.4	PP2A regulatory subunit

* *Saccharomyces* Genome Database name (<http://www.yeastgenome.org/>). Protein names in bold encode previously identified Cdk1 substrates (see main text).

† Cell-cycle-dependent mRNA transcription peak⁸.

‡ In those cases where protein was not detectable, an arbitrary value of 0.1 ng was assigned.

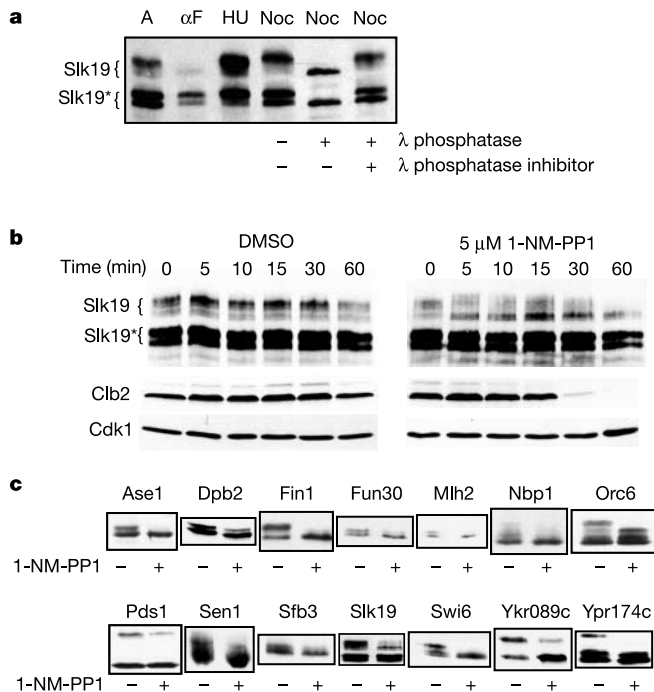


Figure 3 Validation of yeast Cdk1 substrates *in vivo*. **a**, Slk19 is a phosphoprotein with maximal phosphorylation in mitosis. Slk19-Myc₁₃ was immunoprecipitated from a lysate of asynchronous cells (A) or cells arrested in G1 with alpha factor (α F), in S phase with hydroxyurea (HU), or in mitosis with nocodazole (Noc), and immunoblotted with anti-Myc antibodies. Immunoprecipitates in the last two lanes were treated with lambda phosphatase, with or without phosphatase inhibitors. Asterisk indicates separate cleavage products. **b**, Inhibition of Cdk1 results in the rapid dephosphorylation of Slk19. Asynchronous *cdk1-as1 SLK19-MYC₁₃* cells were treated with DMSO or 5 μ M 1-NM-PP1 for the indicated time, lysed in urea buffer, analysed by SDS-PAGE, and immunoblotted for Slk19-Myc₁₃, Clb2 and Cdk1. **c**, Cdk1 inactivation causes a complete or partial collapse in the mobility of 12 new Cdk1 substrates. Asynchronous yeast cultures containing the indicated epitope-tagged protein were grown to mid-log phase, treated with DMSO (–) or 20 μ M 1-NM-PP1 (+) for 15 min at 30 °C, lysed in urea lysis buffer, and analysed by immunoblotting.

molecular targets of Cdk1 have remained unidentified. Mitotic spindle assembly, for example, involves a broad range of molecular motors and other microtubule-associated proteins, but little is known about how these components are regulated. Several of these proteins were identified as Cdk1 substrates in our screen, and it seems likely that the analysis of their regulation by Cdk1 will lead to insights into the control of spindle assembly and stability during mitosis.

The majority of the substrates, including about 50 that are encoded by ORFs with no known function, have not been implicated previously in cell-cycle-dependent events. Studies of these substrates may lead to the discovery of unforeseen regulatory connections in cell-cycle control.

It is now possible to estimate the total number of Cdk1 targets in the proteome. Of the random 198 proteins we tested, sixteen (8.1%) had *P*-scores of 2.0 or greater. If this fraction is representative of the entire yeast proteome, then there may be about 500 Cdk1 substrates in yeast. A number of this magnitude does not seem unreasonable, given that we did not test ~900 ORFs that have a single full Cdk consensus site and thousands more that contain multiple copies of the minimal consensus site. Identification of all Cdk1 substrates will require scanning the entire proteome, either by automated methods with proteomic libraries or by the identification of endogenous Cdk1-as1 targets in cell lysates.

It is possible to generate analogue-sensitive mutant forms of almost any protein kinase, although in some cases multiple mutations are required to produce a mutant that retains normal enzymatic activity^{26,27}. Our method can therefore be used for the systematic identification of substrates for any protein kinase. Such comprehensive analyses of protein kinase targets are clearly an important step towards extracting useful functional information from genome sequence. □

Methods

Yeast methods

All strains were derivatives of W303 or S288C, and were grown at 30 °C unless otherwise noted. Construction of epitope-tagged strains was performed as described²⁸. Ase1 and Slk19 were Myc₁₃-tagged, Nbp1 was TAP-tagged, and Fin1, Orc6 and Pds1 were HA₃-tagged. C-terminally TAP-tagged versions of 169 of the 181 candidate substrates, expressed at their endogenous loci, were obtained from a proteomic library kindly provided by E. K. O'Shea and J. S. Weissman. For western blotting, protein extracts were prepared by bead beating in urea lysis buffer (20 mM Tris-HCl pH 7.4, 7 M urea, 2 M thiourea, 4% CHAPS, 1% DTT, 50 mM NaF, 80 mM beta-glycero-phosphate, 1 mM sodium orthovanadate, 1 mM PMSF). Immunoprecipitation and phosphatase assays were performed as described¹³.

Kinase assays in cell extracts

Cdk1-His₆, Cdk1-as1-His₆, and MBP-Clb2 were purified as described⁶. Active Cdk1-as1-Clb2-TAP was purified from *cdk1-as1 sic1Δ pGAL-CLB2-TAP* yeast cells as described²⁹, except that cell lysates were prepared by bead beating. GST-tagged protein expression was performed as described⁷. Spheroplasting was performed using Zymolyase-100T (Seikagaku America Inc.) in 1.5-ml deep-well 96-well plates. Cells were disrupted by hypotonic lysis with detergent (25 mM Hepes-NaOH pH 7.5, 10 mM NaCl, 1 mM EDTA, 0.1% Triton X-100, 1 mM PMSF, 1 μ g ml^{–1} aprotinin, leupeptin, pepstatin A) and proteins were solubilized by the addition of 300 mM NaCl for 10 min (resulting in efficient extraction of nuclear proteins). Phosphatase inhibitors were not present during lysate preparation, which generally results in dephosphorylation of proteins in the lysate. Lysates were clarified by centrifugation. 500 μ g of whole-cell extract was mixed with 7 nM Cdk1-as1-MBP-Clb2 or 7 nM Cdk1-as1-Clb2-TAP, a creatine-kinase-based 1 mM ATP-regenerating system, 10 × kinase reaction buffer (200 mM Hepes-NaOH pH 7.4, 100 mM NaCl, 10 mM MgCl₂), 10 × phosphatase inhibitors (500 mM NaF, 800 mM β -glycero-phosphate, 10 mM Na₂VO₄), and 5 μ Ci γ -³²P-N⁶-(benzyl) ATP³⁰ and brought to a final volume of 200 μ l. After a 30-min reaction at room temperature, GST-tagged proteins were purified as described previously⁷. Glutathione eluates were loaded onto 9% sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) gels, silver-stained, and exposed to a PhosphorImager cassette (Molecular Dynamics) for 72 h.

A key advantage of our approach is that the kinase reaction is performed in a whole-cell extract, which approximates the normal cellular environment better than typical kinase reactions performed with purified substrates. Protein kinases in crude extracts are exposed to a wide variety of potential protein substrates. If kinase and substrate concentrations are minimized, as in our studies, then only targets with high affinity for the kinase will be phosphorylated, while non-specific phosphorylation events will be reduced.

In the absence of the ATP-regenerating system, the addition of radiolabelled N⁶-(benzyl) ATP leads to the phosphorylation of several proteins in the extract, indicating that some protein kinases are able to use the bulky analogue at a low but appreciable rate (data not shown). This background phosphorylation is presumably lost in the presence of high ATP concentrations (Fig. 1, lane 1) because wild-type kinases are more selective for regular ATP. By contrast, the Cdk1-as1 mutant is much more selective for the radiolabelled bulky analogue, so that its activity with the analogue remains high even in the presence of abundant ATP. It is likely that the addition of ATP also protects the radiolabelled ATP analogue from hydrolysis by the abundant ATPases in the cell extract.

Quantification of phosphorylation, protein levels and *P*-score

For each GST-tagged protein, phosphorylation levels were measured using Imagequant (Molecular Dynamics). Many of the tagged proteins were accompanied by multiple degradation products, which were present before cell lysis, as they were present even when cells were lysed in boiling SDS-PAGE sample buffer (data not shown). For proteins with multiple degradation products, the phosphorylation level represents the sum of the phosphorylation of every degradation product. Phosphorylation levels were normalized between gels and radiolabelled analogue batches by measuring background radioactivity levels for each gel. Protein levels were estimated by comparing protein band intensity, size and shape to BSA standards. *P*-scores were calculated by taking the logarithmic value of the normalized phosphorylation divided by the estimated protein level.

For simplicity, we assumed in these measurements that all silver-stained protein bands beneath the full-length tagged protein are fragments of that protein. However, some of these bands may represent associated proteins. It is remotely possible that these proteins were phosphorylated, resulting in an artefactual increase in the apparent *P*-score. Similarly, fragmentation of tagged proteins could lead to false positives due to exposure of phosphorylation sites that are normally inaccessible.

We suspect that several proteins were not phosphorylated in our screen but are actually Cdk1 substrates. One possible cause of false negatives is that degradation of tagged proteins can result in the loss of phosphorylation sites. Cln2, for example, was detected only as short N-terminal fragments lacking the C-terminal region that contains the Cdk consensus sites. Another source of false negatives is that many of the tagged proteins are

not present in cell lysates in appreciable amounts. Cdc6, for example, was not detectable in cell lysates by silver stain or western blot, suggesting that it is either not expressed in the library or that it is insoluble in our lysis conditions. Because of the method by which the GST-ORF library was constructed, it is likely that each yeast strain contains a mixture of cells expressing either GST alone or the desired GST-tagged protein. As there is no selection for the GST-ORF-containing cells, long-term culture could result in cell lysates containing very low levels of GST-tagged protein and high levels of GST (as illustrated in Fig. 2a, bottom panel). This may explain the very low levels of tagged protein expression we observed (Fig. 2c), and suggests that there are probably more Cdk substrates among the 156 proteins that were not detectable by silver stain and apparently not phosphorylated.

Received 11 March; accepted 19 September 2003; doi:10.1038/nature02062.

- Morgan, D. O. Cyclin-dependent kinases: engines, clocks, and microprocessors. *Annu. Rev. Cell Dev. Biol.* **13**, 261–291 (1997).
- Nigg, E. A. Mitotic kinases as regulators of cell division and its checkpoints. *Nature Rev. Mol. Cell Biol.* **2**, 21–32 (2001).
- Shah, K., Liu, Y., Deirmengian, C. & Shokat, K. M. Engineering unnatural nucleotide specificity for Rous sarcoma virus tyrosine kinase to uniquely label its direct substrates. *Proc. Natl Acad. Sci. USA* **94**, 3565–3570 (1997).
- Shah, K. & Shokat, K. M. A chemical genetic screen for direct v-Src substrates reveals ordered assembly of a retrograde signaling pathway. *Chem. Biol.* **9**, 35–47 (2002).
- Witucki, L. A. *et al.* Mutant tyrosine kinases with unnatural nucleotide specificity retain the structure and phospho-acceptor specificity of the wild-type enzyme. *Chem. Biol.* **9**, 25–33 (2002).
- Bishop, A. C. *et al.* A chemical switch for inhibitor-sensitive alleles of any protein kinase. *Nature* **407**, 395–401 (2000).
- Martzen, M. R. *et al.* A biochemical genomics approach for identifying genes by the activity of their products. *Science* **286**, 1153–1155 (1999).
- Spellman, P. T. *et al.* Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol. Biol. Cell* **9**, 3273–3297 (1998).
- Moll, T., Tebb, G., Surana, U., Robitsch, H. & Nasmyth, K. The role of phosphorylation and the CDC28 protein kinase in the cell cycle-regulated nuclear import of the *S. cerevisiae* transcription factor SWI5. *Cell* **66**, 743–758 (1991).
- Verma, R. *et al.* Phosphorylation of Sic1p by G1 Cdk required for its degradation and entry into S phase. *Science* **278**, 455–460 (1997).
- Deshais, R., Chau, V. & Kirschner, M. Ubiquitination of the G1 cyclin Cln2p by a Cdc34p-dependent pathway. *EMBO J.* **14**, 303–312 (1995).
- Zachariae, W., Schwab, M., Nasmyth, K. & Seufert, W. Control of cyclin ubiquitination by CDK-regulated binding of Hct1 to the Anaphase Promoting Complex. *Science* **282**, 1721–1724 (1998).
- Jaspersen, S. L., Charles, J. F. & Morgan, D. O. Inhibitory phosphorylation of the APC regulator Hct1 is controlled by the kinase Cdc28 and the phosphatase Cdc14. *Curr. Biol.* **9**, 227–236 (1999).
- Gartner, A. *et al.* Pheromone-dependent G1 cell cycle arrest requires Far1 phosphorylation, but may not involve inhibition of Cdc28-Cln2 kinase, *in vivo*. *Mol. Cell Biol.* **18**, 3681–3691 (1998).
- Mortensen, E. M., McDonald, H., Yates, J. III & Kellogg, D. R. Cell cycle-dependent assembly of a Gin4-septin complex. *Mol. Biol. Cell* **13**, 2091–2105 (2002).
- McMillan, J. N., Theesfeld, C. L., Harrison, J. C., Bardes, E. S. & Lew, D. J. Determinants of Swe1p degradation in *Saccharomyces cerevisiae*. *Mol. Biol. Cell* **13**, 3560–3575 (2002).
- Elsasser, S., Chi, Y., Yang, P. & Campbell, J. L. Phosphorylation controls timing of Cdc6p destruction: A biochemical analysis. *Mol. Biol. Cell* **10**, 3263–3277 (1999).
- Nguyen, V. Q., Co, C. & Li, J. J. Cyclin-dependent kinases prevent DNA re-replication through multiple mechanisms. *Nature* **411**, 1068–1073 (2001).
- Masumoto, H., Muramatsu, S., Kamimura, Y. & Araki, H. S-Cdk-dependent phosphorylation of Sld2 essential for chromosomal DNA replication in budding yeast. *Nature* **415**, 651–655 (2002).
- Agarwal, R. & Cohen-Fix, O. Phosphorylation of the mitotic regulator Pds1/securin by Cdc28 is required for efficient nuclear localization of Esp1/separase. *Genes Dev.* **16**, 1371–1382 (2002).
- Rudner, A. D. & Murray, A. W. Phosphorylation by Cdc28 activates the Cdc20-dependent activity of the anaphase-promoting complex. *J. Cell Biol.* **149**, 1377–1390 (2000).
- Roberts, J. M. Evolving ideas about cyclins. *Cell* **98**, 129–132 (1999).
- Cross, F. R., Yuste-Rojas, M., Gray, S. & Jacobson, M. D. Specialization and targeting of B-type cyclins. *Mol. Cell* **4**, 11–19 (1999).
- Sullivan, M., Lehane, C. & Uhlmann, F. Orchestrating anaphase and mitotic exit: separase cleavage and localization of Slk19. *Nature Cell Biol.* **3**, 771–777 (2001).
- Ficarro, S. B. *et al.* Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*. *Nature Biotechnol.* **20**, 301–305 (2002).
- Niswender, C. M. *et al.* Protein engineering of protein kinase A catalytic subunits results in the acquisition of novel inhibitor sensitivity. *J. Biol. Chem.* **277**, 28916–28922 (2002).
- Habelhah, H. *et al.* Identification of new JNK substrate using ATP pocket mutant JNK and a corresponding ATP analogue. *J. Biol. Chem.* **276**, 18090–18095 (2001).
- Longtine, M. S. *et al.* Additional modules for versatile and economical PCR-based gene deletion and modification in *Saccharomyces cerevisiae*. *Yeast* **14**, 953–961 (1998).
- Puig, O. *et al.* The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods* **24**, 218–229 (2001).
- Kraybill, B. C., Elkin, L. L., Blethrow, J. D., Morgan, D. O. & Shokat, K. M. Inhibitor scaffolds as new allele specific kinase substrates. *J. Am. Chem. Soc.* **124**, 12118–12128 (2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank E. Phizicky for the GST-ORF library; the many laboratories at UCSF who helped duplicate and distribute the GST-ORF library; E. K. O'Shea and J. S. Weissman for TAP-tagged strains; D. Kellogg for sharing unpublished results; E. K. O'Shea and R. Deshaies for expression plasmids; N. Dephoure and M. Verzi for strain construction; S. Biggins, D. Kellogg, A. Goga, S. Steggerda, C. Carroll, D. Randle, M. Enquist-Newman and M. Loog for advice and critical reading of the manuscript. This work was supported by funding from the National Institute of General Medical Sciences (D.O.M.), the National Institutes of Health (K.M.S.), and by a predoctoral fellowship from the National Science Foundation (J.A.U.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.O.M. (dmorgan@cgl.ucsf.edu).

Seeing is believing

Peering down an eyepiece is becoming a thing of the past. Tim Chapman takes a look into the digital world of a new generation of microscopes and imaging systems.

If anything has remained constant over centuries of biological research, it is the value of the evidence of your own eyes, gained by examining specimens — living or dead — in as much detail as possible. Today's microscopes are, of course, a world away from the optical systems used by the pioneers of biology. Researchers in genomics and proteomics need to study molecular interactions within living cells in detail, whereas the high-throughput screening required by the pharmaceutical industry demands the automated scanning and analysis of large numbers of samples.

Going digital

Even for the simplest applications, microscopes have embraced digital technology for image analysis and, increasingly, for presenting real-time images without the operator needing to peer down eyepieces. "No company is selling a lot of film-based cameras any more," says Rainer Wegerhoff, section manager at the microscopy training facility run by microscope manufacturers Olympus in Hamburg, Germany. "Everything is going into the digital market because there is so much more benefit for users to use digital

imaging instead of film-based imaging."

Digital imaging can be particularly helpful in bioscience applications involving fluorescence. "If you see fluorescence, it's very hard to get good quantifiable image data from a film-based system," Wegerhoff notes. "Most people would start with digital imaging systems. You have only a few milliseconds, so you can resolve single-molecule detection or things that were not possible with video or film."

Although most digital microscopes resemble traditional optical systems with an added facility to display the image digitally, there is now a new generation of innovative microscopes for exclusively digital use, which take the image directly to a computer screen. Olympus's MIC-D is a relatively simple low-power digital microscope designed for education and basic laboratory applications. It has no eyepiece, but plugs into a desktop or laptop computer through a USB connection. Its base houses a low-power CMOS (complementary metal oxide semiconductor) detector, while a swivelling lamp arm with a solid-state LED lamp allows the sample to be illuminated from a range of angles, and in various illumination modes.

The CMOS detector produces an image of just 640×480 pixels, less than a tenth of many laboratory systems, and can be used to look, for example, at pond life and other small creatures, and the general cellular structure of fixed plant tissues. "It's not comparable to real research microscopes; it's an instrument that can be transported from one location to another and used in the field," Wegerhoff says. Several police forces have shown interest in the MIC-D for initial forensic imaging at crime scenes.

Many applications do require higher resolutions, particularly in the medical market. Pathologists, who have traditionally made diagnoses from images recorded on silver halide film, are beginning to use digital images now that charge-coupled devices (CCDs) offering similar resolution to film are more widely available.

The Coolscope system from Nikon, based in Kanagawa, Japan, is specifically targeted at the pathology market. It is based on a 5-megapixel CCD, similar to that used in Nikon's general-purpose DS-5M digital microscope. But the Coolscope combines the microscope, CCD camera and network functions into an integrated tower unit that

LIGHTING UP THE BODY

The ability to track the molecular action of disease or other biological processes within live subjects is a potentially invaluable tool for studying the broad array of drug targets identified by genomic and proteomic research.

Medical imaging techniques such as magnetic resonance imaging (MRI) and positron emission tomography (PET) can be deployed, but they demand expensive equipment beyond the reach of many labs and provide limited functional information. A simple but effective alternative is to use optical imaging techniques that exploit the properties of light-emitting enzymes.

This approach was pioneered at Stanford University in the mid-1990s by husband-and-wife team Christopher and Pamela Contag, together with David Benaron. The Stanford team infected mice with bacterial pathogens that had been engineered to carry the gene for luciferase, the enzyme that makes fireflies glow. As these pathogens spread through the mice, light emitted by the luciferase could be detected with a highly sensitive CCD camera.

In 1997, Pamela Contag left Stanford to form the company Xenogen to commercialize the technology. Based in Alameda, California, Xenogen now produces both the sensitive light recorders and software required for what it calls biophotonic imaging, and a family of transgenic mice and rats



Xenogen's Pamela Contag.

engineered to emit light when tagged genes are activated.

Because all imaging is *in vivo*, the technology allows longitudinal study design. "You start and end the experiment with the same animal and monitor the same animal over time — this allows you to use 60–80% fewer animals," Contag notes. "That allows experiments to be run at more reasonable cost, and because we're looking at very early in the disease process, these disease models are less stressful for the animals."

The company has now signed licensing deals with a string of pharmaceutical companies including AstraZeneca and Bristol-Myers Squibb. "More and more we're using our technology for drug discovery," says Contag. "For instance, we have a whole series of animals that respond to toxicological stress. Or if drug metabolism is an issue, we have a series of animals that will tell you which enzyme is turned on."

Xenogen is now developing its technology for use with human subjects, and is also working on improving image quality. Although current resolutions cannot match those of PET or MRI, the technique can provide much more functional information. "There are things you can learn using optical imaging you cannot using PET and MRI," Contag says. "This is a desktop instrument, and that makes a huge difference as to where you apply it. It's very easy to use and the results are pretty spectacular."

T.C.

bears little resemblance to the traditional microscope. The user just has to add a monitor and a computer mouse.

"The Coolscope is basically a bright-field imaging system," says Robert Forster, general manager for Nikon Instruments UK. "It's primarily aimed at the clinical market for consultant pathologists to share images in multidisciplinary meetings with surgeons and radiologists. The Coolscope enables them to look at a patient's specimen on a microscope glass slide. It also enables them to view that slide remotely and to control the functions from anywhere in the world, providing you get the correct access through the hospital networks."

Digital technology is changing lab microscopes in many different ways. The DM DigitalMicroscope range produced by Leica Microsystems of Wetzlar, Germany, is not based on a digital camera as the name might imply, although Leica's DC range of cameras can be added as extras. Here, 'digital' refers to an integrated user-interface and control system.

"What scientists want is to concentrate on their work," says Michael Ganser, project leader for the Leica DM4000 B and DM5000 B microscopes for biological applications. "Ten or fifteen years ago, scientists loved their microscopes and knew exactly how to work them, but this has changed and microscopes are becoming more and more of a tool. What we tried to do is make it really easy to operate the microscope."

The DM series includes what Leica calls

"intelligent automation" — push-button controls to rapidly configure the microscope for complex procedures. Automatic aperture and light adjustment, and colour-intensity control allow the microscope to keep itself at the optimum settings. The DM4000 B and DM5000 B are mainly aimed at biomedical applications and pathology, and can be integrated with control systems for fluorescence, cytogenetics or other common techniques.

Live cell imaging

With the completion of the human genome sequence, the emphasis in biomedical research is now firmly back on functional biology and attempts to characterize and understand the proteome. To do this, researchers want to look inside live cells to study the expression and function of genes and proteins within particular pathways.

Using fluorescence to image living cells has become a major application for microscope producers and users. As well as its stand-alone microscopes, Leica Microsystems also produces the AS MDW, a dedicated workstation for the rapid imaging of molecular interactions in live cells. "All the hardware components, such as camera and piezo-focus, are designed in a way to optimize the speed of acquiring images," says Werner Kampe, marketing manager for Leica's compound-microscopy business unit. "That's important not just for scientists; it's also important not to damage the living cells under the microscope. The



Easy viewing: Nikon's Coolscope.

longer they're under the light, the shorter their lifetime is." Sophisticated software means that cell interactions can be recorded in three dimensions, and merged into a time-lapse movie of changes over several days — something that Leica refers to as four-dimensional imaging.

A rival system from PerkinElmer of Wellesley, Massachusetts, boasts six-dimensional imaging of live-cell interactions. As well as three dimensions of space and one of time, the UltraView RS can image cellular interactions at two additional wavelengths. "To be able to do that for a live cell and get meaningful data you need to be sampling every few seconds," says Terry McCann, cellular sciences business manager at PerkinElmer. "UltraView RS takes advantage of the speed of the system so that we can generate these multidimensional images on biologically relevant timeframes."

EASING THE STRAIN

Many microscope users know all too well that prolonged use can be a real pain. Many systems demand that the user maintains a rigid body posture which can become increasingly uncomfortable, and concentrating on the magnified image for long periods can lead to eye fatigue and nausea. It can all add up to a serious reduction in accuracy and efficiency in the short term, and a long-term risk of debilitating repetitive strain injury (RSI).

Many manufacturers now design their microscopes to be ergonomically efficient, making operation as simple and intuitive as possible while

minimizing the risk of strain or fatigue. Leica Microsystems of Wetzlar in Germany, for example, worked with the Fraunhofer Institute of Stuttgart, Germany, to design its DM series of microscopes. The main control panels are positioned for direct and intuitive access, and height, viewing angle and eyepiece length can be adjusted to suit the user.

The narrow field of view from most microscope eyepieces is a major cause of eye strain and bad posture. Users who wear spectacles often have to remove them, increasing the risk of eye strain; and many users also suffer the distraction of floating fragments of tissue debris in the eye.

Vision Engineering of Woking, UK, a company specializing in microscope ergonomics, claims to have solved these problems with an eyepiece called the Isis expanded-pupil system. Fitted in place of the traditional eyepiece, the Isis is based on a small disc containing more than 2 million microlenses. This disc rotates at about 5,000 r.p.m., so that the millions of individual optical paths are merged to form a single high-clarity image, with an exit field around ten times as wide as from a conventional eyepiece.

"The key thing is the image, and where the image leaves the microscope and meets the operator," says international marketing manager Geoff Collins. "Our technology allows the image to have a much wider exit pupil, which means you can see it from much further away, from a greater range of angles. That allows you to move your head around, wear spectacles, and be further away from the eyepiece." Isis is available on Vision's own range of laboratory microscopes, and as an accessory to replace the conventional eyepieces on a wide range of commercial microscopes.

T.C.



Isis eyepiece from Vision Engineering gives a wider view.

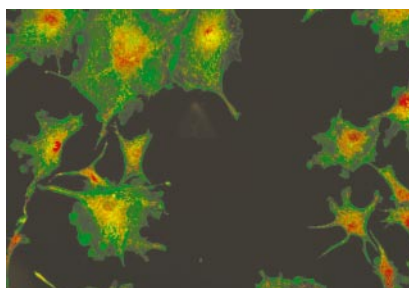
The UltraView system is based on confocal microscopy, an optical sectioning technique that uses a scanning laser to produce high-resolution images from the fluorescence emitted by the sample. Traditional confocal microscopes can be too slow to study the most rapid cell processes, and the intensity of laser light can damage living cells.

Rather than using a single pinpoint laser beam, UltraView uses a broad beam imaged through two rapidly rotating discs: one, called a Nipkow disc, contains several thousand pinpoint holes; the other is an aligned disc featuring an equal number of micro-lenses to focus light onto the former. "You end up with 1,000 laser beams which shower the sample with light," says McCann. "Instead of one point being imaged, you have 1,000 points being imaged. You can see the whole field of view more quickly, and you also use very much less light."

High-throughput scanning

Live-cell imaging is also being taken into the high-throughput environment of industrial drug screening. The combination of high-throughput screening and high-content imaging of cells was pioneered by Pittsburgh-based Cellomics (see 'High throughput goes 3D'), and the major producers of automated lab equipment are now moving into the field.

The IN Cell Analyzer 3000 produced by Amersham Biosciences of Buckinghamshire, UK, is a confocal imaging system designed to perform high-resolution cellular



ImageXpress: neural stem cells.

assays at the rates demanded by today's automated drug-discovery labs.

"There's an increasing need to understand what's going on within the cell, and to do that in the environment of high-throughput screening is extremely valuable," says John Anson, vice-president for bioassay development at Amersham. "The objectives and a lot of the resolution come from existing technology, but what we've been able to do is build in the robustness and speed to take this through to the high-throughput environment."

Although the system uses laser-based autofocus and other technologies to ensure rapid imaging of its well plates, the key to high throughput is in the software. "What we've been able to do is develop a software regime with the capability and speed not only to capture the data but to analyse it in a very short space of time," Anson says. "In the time it takes to mechanically move from one cell to the next, the

software is doing all the work to generate knowledge from the image it's just taken." Integrated software also allows data to be combined with data from other Amersham systems such as its Typhoon and Storm series of gel scanners. "This technology isn't necessarily stand-alone," notes Anson. "Everyone would be interested in comparing data from cellular assays with *in vitro* assays such as gene expression in gel."

Integrated software is also at the heart of the high-throughput capabilities of the ImageXpress system produced by Axon Instruments of Union City, California. Like Amersham's IN Cell, ImageXpress is a rapid, robust automated system for imaging and analysing live cells, with applications in target evaluation, compound optimization and toxicology tests. The system incorporates a general-purpose analysis engine that can be applied to many different assays through a toolbox of dedicated software. Axon has recently developed sophisticated neuronal image-analysis software in association with the Australian research body CSIRO, which is now included in ImageXpress's toolbox.

Automated operation and image processing are based on the common VB scripting protocol, allowing ImageXpress to be integrated with plate-loading robots and other automated systems. Axon's Acuity image-analysis software, meanwhile, allows the import of data from other systems for multiparametric data mining.

"I think this technology is reaching a consolidation phase where the capabilities of the

HIGH THROUGHPUT GOES 3D

High-throughput analysis of cellular processes is now entering the mainstream for drug discovery, but Pittsburgh-based Cellomics has been working in this area for some time. Founded in 1996 to focus on the automated analysis of cell arrays, Cellomics offers an integrated package of imaging platforms with full environmental control, bioassays and bioinformatics and analysis software.

The firm produces two core platforms: KineticScan analyses cellular and intracellular interactions over time, and can deal with either live or fixed cells, whereas ArrayScan generates information-rich data on the effects of drug candidates on cells and is aimed at target validation and lead optimization applications.

The modular architecture of ArrayScan VTI, the fifth and latest generation of the system, allows users to increase optical capabilities by adding filters and objectives, and software to handle and visualize data better. The system also addresses the two conflicting aims of array screening — the need for high throughput, and the need for information-rich, high-resolution data including three-dimensional imaging.

"It takes time to acquire various z-positions and use higher magnification, so throughput is at odds with 3D resolution," says Martin Pietila, high-content screening instrumentation product manager at Cellomics. "For the first time we've been able to offer both operating modes."

The latest ArrayScan includes the ApoTome optical-sectioning device developed by Carl Zeiss Microscopy of Jena, Germany, which can be deployed as required. ApoTome allows confocal-style three-dimensional

imaging with a conventional fluorescence microscope, by projecting the image of a grid structure into the focal plane of the specimen. Combining images with the grid in three positions within this plane produces an optical section with improved contrast and resolution in the z-axis.

"This affords scientists precise control of sampling applications," Pietila says. "It allows them to take multiple measurements of different z-positions to look for structures with variable distribution in three dimensions." The ArrayScan VTI can analyse up to 150 96-well plates a day using ApoTome, or 260 a day without.

T.C.



ArrayScan VTI: high-throughput screening goes confocal.

systems are starting to be appreciated and people are looking at what can be done with large amounts of data from these kinds of instruments — at multiparametric studies of cells and what kinds of coordinations you can derive between different families of cells,” says Andrew Olson, product line manager for high-throughput cellular imaging at Axon.

Electron microscopes

Cell biologists are also turning to electron microscopy to understand supramolecular structures at the atomic level. Transmission electron microscopy (TEM) has long been used to visualize cell ultrastructure in plastic-embedded sections of biological material, but it is now increasingly being used for three-dimensional imaging of whole cells, cell organelles and protein complexes.

To understand how protein complexes execute biological functions, researchers need to know the structures of these large, fragile constructions, and have deployed techniques such as nuclear magnetic resonance (NMR) spectroscopy, which offers only limited resolution, and X-ray crystallography, which requires large numbers of molecules in a crystalline array. TEM offers a more flexible solution.

“Over the past five years, the proof of concept has been given that TEM can be very instrumental in determining the three-dimensional structure of these huge protein complexes,” says Werner Hax, life-sciences business development manager at electron-



Tecnai G2 Polara from FEI.

microscope manufacturer FEI of Hillsboro, Oregon. “TEM can tell you exactly where in the cell these proteins and protein complexes are located, so you can bring everything into perspective.”

The Tecnai G2 Polara from FEI is a TEM optimized for structural biology tomography and single-particle analysis. “Imaging of frozen hydrated biological samples within a TEM has to be done with the greatest care, as the electron beam can damage the object under investigation,” says Hax. “The Tecnai G2 Polara allows observation of proteins at temperatures below 10 K and provides the other conditions to minimize beam damage.”

The Tecnai instruments are fully digitally controlled, with accessories such as cameras and X-ray detectors embedded, allowing them to be integrated into auto-

mated processes. “A lot of data have to be acquired and a lot of computational technology has to be applied to reconstruct the 3D representation of each protein in the cell,” Hax says. “To do that within an acceptable period of time, particularly if you want to get this applied in industry in drug discovery, automation is an absolute must. Without that, TEM will never have the dominant role it can play.”

Less specialized electron microscopes are finding wider use as an everyday tool for biological research. Delong Instruments of Brno, Czech Republic, has produced what it claims is the world's smallest TEM. The LVEM5 is intended primarily for biological and medical research, and provides high-contrast images of thin sections and particles such as viruses, enzymes, ribosomes, proteins and DNA, without the need for staining with heavy metals.

“This is practicable only by a considerable decrease of the energy of the imaging beam electrons,” says Delong's Michal Drsticka. “When the energy is decreased roughly ten times, the contrast is ten times higher. The microscope therefore can be used both in university research and biological objects diagnostics.”

The LVEM5 is around a tenth of the size of a conventional TEM, and could join the optical microscope as an everyday benchtop instrument. The bioscience researcher has never had such a choice of ways of seeing. ■

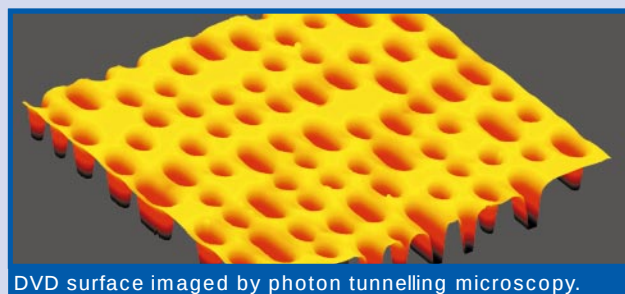
Tim Chapman is a freelance technology journalist based in Halifax, UK.

SCANNING THE SURFACE

Photon tunnelling microscopy (PTM) is a unique surface-scanning technology that can produce real-time three-dimensional images at unparalleled resolution of biological samples, including live cells. The basic technology can be used with common optical microscopes, and can achieve better vertical resolution than a scanning electron microscope without the need for a scanning tip.

The system addresses a fundamental problem of optical imaging — the smaller an object is relative to the wavelength of incident light, the larger the angle of the diffracted light. For very small objects, the light is diffracted at such a large angle that it cannot leave the object, and is bound in an ‘evanescent field’, the intensity of which decays exponentially with distance. To see such a small object, you have to take the lens to within a few hundred nanometres of the object's surface without damaging either the instrument or the sample. And unless you operate in a ‘clean room’, you will have to deal with dust that is many times larger than this distance.

The technology that enables PTM was developed by John Guerra in his former role as senior principal engineer at Polaroid. In June 2002, Guerra launched his own company, Nanoptek, based in Concord, Massachusetts, to develop and commercialize the technology. The key to Nanoptek's approach is a square of soft, flexible polymer film, 75 mm across and 15–20 μm thick, called a transducer, which is applied over the sample. This is made optically continuous to the near-field lens with a fluid coupling. “This allows you to move around this 75-mm area very quickly without damaging the microscope or the sample,” Guerra explains. The film also covers any dust particles on the sample.



DVD surface imaged by photon tunnelling microscopy.

Over the 400-nm vertical range of the evanescent field, the coupling goes from total to zero, creating a grey-scale image in the microscope. A 12-bit CCD camera would give a vertical resolution of 0.1 nm — better than that achievable with scanning electron microscopy (SEM), and equal to that of atomic force microscopy (AFM), with the added advantage of real-time imaging.

“You can get 200- μm fields of view at video rates. You can watch chemotaxis in progress or cells moving,” says Guerra. At about 100 nm, lateral resolution is not as good as with SEM or AFM, but is better than that of confocal or phase-shifted interference microscopy. Nanoptek is also planning to combine PTM with phase-shifting to increase the lateral resolution to 0.1 nm. “That will be the next exciting development,” Guerra says. “You will be able to image viruses and very, very small biological materials.”

T.C.

Company	Products/activity	Location	URL
Optical and fluorescence microscopy			
Applied Scientific Instrumentation	Automated systems for microscopy and fluorescence screening	Eugene, Oregon	www.asiimaging.com
Asylum Research	Molecular Force Probe 1D and 3D atomic-force microscopes	Santa Barbara, California	www.asylumresearch.com
Biomedical Photometrics	TISSUEscope confocal laser scanning system for fluorescent, transmitted and reflected light scanning of biological tissue samples	Waterloo, Ontario	www.confocal.com
Cambio	STAR FISH chromosome paints	Cambridge, UK	www.cambio.co.uk
Carl Zeiss	ApoTome imaging system; PALM laser microdissection system; automated systems for ultra-high-throughput screening	Jena, Germany	www.zeiss.de
Cell Robotics International	Microscope workstations for biomedical research	Albuquerque, New Mexico	www.cellrobotics.com
Chroma Technology	Optical filters and fluorescence filter sets	Brattleboro, Vermont	www.chroma.com
Energy Beam Sciences	Products for light microscopy and electron microscopy	Agawam, Massachusetts	www.ebsciences.com
Hamamatsu Photonic Systems	Digital imaging systems and cameras	Bridgewater, New Jersey	usa.hamamatsu.com
Imaging Research	MCID fluorescence microscopy system; image-analysis software; microarray analysis and statistical validation software	St Catharines, Ontario,	www.imagingresearch.com
Intracellular Imaging	Digital fluorescence imaging, microphotometry and epifluorescence microscopes	Cincinnati, Ohio	www.intracellular.com
JPK Instruments	NanoWizard atomic force microscope	Berlin, Germany	www.jpk.com
Leica Microsystems	DM digital microscope range, optical microscopes and microscope systems	Wetzlar, Germany	www.leica-microsystems.com
Lighttools	Whole mouse imaging, fluorescent imaging gel documentation systems, thermal paper	Encinitas, California	www.lighttools.com
Mauna Kea Technologies	Cell-viZio fluorescence imaging system for <i>in vivo</i> and <i>in situ</i> microscopy	Paris, France	www.maunakeatech.com
Micro Video Instruments	Microscopes and accessories	Avon, Massachusetts	www.mvi-inc.com
Molecular Imaging	Developer and supplier of scanning-probe and atomic-force microscope systems, accessories and software	Tempe, Arizona	www.molec.com
Vision Engineering	Microscope manufacturers; microscope accessories	Woking, UK	www.visioneng.com
Nanoptek	Digital photon tunnelling microscope	Concord, Massachusetts	www.nanoptek.com
Nikon Instruments	Microscopes, COOLSCOPE digital microscope with screen	Melville, New York	www.nikonusa.com
Olympus America	MIC-D digital microscope, upright and inverted microscopes, stereo microscopes, dissecting microscopes, Fluoview confocal laser scanning microscopes	Melville, New York	www.olympus.com
Olympus Europe	MIC-D digital microscope, upright and inverted microscopes, stereo microscopes, dissecting microscopes, Fluoview confocal laser scanning microscopes	Hamburg, Germany	www.olympus-europa.com
Optronics	MicroFire digital microscope imaging cameras	Goleta, California	www.optronics.com
PerkinElmer Life Sciences	UltraView RS confocal laser scanning microscope system for live cell imaging	Boston, Massachusetts	lifesciences.perkinelmer.com
Photon Technology International	Imagemaster fluorescence and general imaging systems	Lawrenceville, New Jersey	www.pti-nj.com
Richardson Technologies	RTM 3.0 microscope system for live-cell imaging; high-resolution field microscope	Bolton, Ontario	www.richardson-tech.com
Tritech Research	Dissecting stereomicroscope system	Los Angeles, California	www.tritechresearch.com
Unitron	Optical microscopes	Bohemia, New York	www.unitronusa.com
Veeco	Scanning-probe microscopes	Woodbury, New York	www.veeco.com
WITec	Optical and scanning probe microscopes	Ulm, Germany	www.witec.de
Xenogen	Luciferase-based <i>in vivo</i> imaging for living animals using CCD-based detection	Alameda, California	www.xenogen.com
Electron microscopy			
Delong Instruments	LVEM5 low-voltage electron microscope	Brno, Czech Republic	www.dicomps.com
FEI Company	Tecnai G2 transmission electron microscopes; scanning electron microscopes; ion-beam technology	Hillsboro, Oregon	www.feicompany.com
Gatan	Instrumentation and software for enhancing the operation and performance of electron microscopes	Pleasanton, California	www.gatan.com
Hitachi High Technologies	Scanning and transmission electron microscopes	Schaumburg, Illinois	www.hitachi-hita.com
Jeol	Scanning and transmission electron microscopes	Peabody, Massachusetts	www.jeol.com
LEO electron microscopy	Scanning and transmission electron microscopes	Thornwood, New York	www.leo-usa.com
Thomson Scientific Instruments	WINEDS PC-based X-ray analyzer for electron microscopes	Carlton, Australia	www.tsi.com.au
4pi	X-ray microanalysis systems, components, and software and digital imaging systems for electron microscopes	Durham, North Carolina	www.4pi.com
Scanners and imaging systems			
Andor Technology	CCD and intensified CCD camera systems	Belfast, UK	www.andor-tech.com
Affymetrix	GeneChip microarray systems and scanner	Santa Clara, California	www.affymetrix.com
Agilent	Catalogue and custom DNA microarrays, scanner and software	Palo Alto, California	www.agilent.com
Alpha Innotech	Alpha Array microarray reader	San Leandro, California	www.alphainnotech.com
Amersham Biosciences	IN Cell Analyzer 3000 confocal imaging system for cell screening; microarray and gel scanners and analysis software	Chalfont St Giles, UK	www.amershambiosciences.com
Applied Precision	arrayWoRx biochip reader	Issaquah, Washington	www.appliedprecision.com
Asys Hitech	Microplate readers, washers and dispensers	Eugendorf, Austria	www.asyshitech.com
Axon Instruments	ImageXpress imaging system for cell screening; GenePix 4000B microarray scanner and GenePix Pro 4.0 microarray analysis software	Foster City, California	www.axon.com

Company	Products/activity	Location	URL
Beckman Coulter	Automated tools for molecular biology, biochemistry, genomics and proteomics	Fullerton, California	www.beckmancoulter.com
Bio-Rad	Molecular Imager FX Pro Plus Multimager System laser-based fluorescence imaging system; CellMap laser scanning confocal system	Hercules, California	www.bio-rad.com
BMG Labtechnologies	Microplate readers and handling systems	Offenburg, Germany	www.bmg-labtechnologies.com
Cellomics	ArrayScan HCS and KineticScan systems for automated cell-based high-content screening for drug discovery	Pittsburgh, Pennsylvania	www.cellomics.com
CyBio	Pipetting, liquid handling, incubation and imaging systems for automated screening	Jena, Germany	www.cybio-ag.com
Dynex Technologies	Absorbance and luminescence microplate readers, microplate washers and automated workstations	Chantilly, Virginia	www.dynextechnologies.com
Fujifilm	Imaging systems for protein electrophoresis and general array analysis	Tokyo, Japan	www.fujifilm.co.jp
GeneFocus	Microscope/MACROscope laser confocal epifluorescence scanning system; DNAscope confocal laser imaging microarray scanners	Waterloo, Ontario	www.genefocus.com
Genetix	Q-bot arrayers, scanners and consumables	New Milton, UK	www.genetix.co.uk
Graffinity	Proprietary drug-discovery platform based on chemical microarrays and label-free imaging of protein–ligand interactions	Heidelberg, Germany	www.graffinity.com
Kodak Digital Imaging	Imaging systems and image-analysis software	Rochester, New York	www.kodak.com
LaVision Biotech	BioAnalyzer biochip readers; TriMScope multifocal multiphoton laser scanning microscope	Bielefeld, Germany	www.lavisionbiotec.de
LI-COR Biosciences	Odyssey Infrared imaging system for western blot and in-cell western blot assay analysis	Lincoln, Nebraska	www.licor.com
Metrigenix	4D Assay scanner for DNA microarrays	Gaithersburg, Maryland	www.metrigenix.com
MiraBio (Hitachi Genetic Systems)	Microarray systems, bead-based liquid array system (SPBIO II spotter and CRBIO lie scanner), software and consumables	Alameda, California	www.mirabio.com
Molecular Devices	Liquid-handling and microplate-processing equipment, imaging instruments and assay systems for integration into robotic systems	Sunnyvale, California	www.moleculardevices.com
NextGen Sciences	a2DE automated 2D gel electrophoresis workstation	Huntingdon, UK	www.nextgensciences.com
Packard Bioscience	Arrayers, scanner and data analysis software	Meriden, Connecticut	www.packardbioscience.com
PCO imaging	Digital video camera systems	Kelheim, Germany	www.pco.de
Polaroid	Digital imaging systems for microscopes	Waltham, Massachusetts	www.polaroid.com
QImaging	CCD digital cameras for imaging in life-science applications	Burnaby, British Columbia	www.qimaging.com
Relab	Arrays, scanners, workstations and dispensers for biochips	Recklinghausen, Germany	www.relab.de
Shimadzu North America	AIM 8800 automated Fourier transform infrared microscope	Columbia, Maryland	www.ssi.shimadzu.com
Tecan	Robotic laboratory workstations and automated solutions for genomics, proteomics; LS series laser scanner	Medford, Massachusetts	www.tecan.com
Thermo Electron	Infrared and Raman microscopes; laser sources; image-analysis software	Waltham, Massachusetts	www.thermo.com/eThermo
TissueInformatics	TissueAnalytics high-throughput automated histology laboratory service, analytical software; drug target discovery	Pittsburgh, Pennsylvania	www.tissueinformatics.com
UVP	Biolmaging systems for visualization, documentation and analysis of gels, membranes and microplates	Upland, California	www.uvp.com
Vysis	Microarrays, microarray readers; fluorescent probes	Downer's Grove Illinois	www.vysis.com

Imaging and image-analysis software

Advanced Research Instruments	AutoSEM 1 image/X-ray analyser for scanning electron microscopy	Wheat Ridge, Colorado	www.aricorp.com
Advanced American Biotechnology & Imaging	Gel image-analysis software	Fullerton California	www.aabi.com
BioSystematica	Gel-imaging systems and image-analysis software	Tavistock, UK	www.biosystematica.com
Clemex Technologies	Image-analysis systems and image-analysis software for quality control and research; image-analysis solutions for quantitative microscopy	Longueuil, Quebec	www.clemex.com
Data Translation	GLOBAL LAB Image/2 image-analysis software	Marlboro, Massachusetts	www.datx.com
DECODON	Software for 2D-gel image-analysis and information storage	Greifswald, Germany	www.decodon.de
GeneBio	Melanie 2D gel image-analysis software	Geneva, Switzerland	www.genebio.com
I-CUBE	Suppliers of image-analysis software	Millersville, Maryland	www.i-cubeinc.com
Media Cybernetics	Image-Pro image-analysis software; Array-ProAnalyzer for microarrays; Gel-ProAnalyzer for gels; IQbase image informatics software; Optimas image-analysis software	Carlsbad, California	www.mediacy.com
Mirero	GAIA image-analysis software	Seoul, Korea	www.gaia-zone.com
Nonlinear Dynamics	Phoretix image-analysis and database software for 1D- and 2D-gel electrophoresis and high-throughput arrays; Progenesis for the analysis of high-throughput 2D-gels for proteomics	Newcastle-upon-Tyne, UK	www.nonlinear.com
Proteomic Solutions	Melanie software for 2D-gel analysis; distributor for Genomics Solutions; proteomics services	St Marcel, France	www.proteomicsolutions.fr
Scanalytics	IPLab and IPLab extensions for microscopy; gel analysis software; microarray software and elispot screening systems; deconvolution and 3D restoration software	Fairfax, Virginia	www.scanalytics.com
Scimagix	SIMS scientific image management systems	San Mateo, California	www.scimagix.com
Scion Corporation	Digital imaging systems and imaging software for microscopy	Frederick, Maryland	www.scioncorp.com
Soft Imaging System	Digital cameras for light and electron microscopes; image-analysis software	Münster, Germany	www.soft-imaging.de
Synoptics	Digital imaging systems for microscopy; gel image and analysis software	Cambridge, UK	www.synoptics.co.uk
VayTek	Digital imaging systems; deconvolution software, 3D volume visualization, measurement and presentation software; digital control for microscope peripherals	Fairfax, Iowa	www.vaytek.com

● see advertisement

Contacts

Publisher: Ben Crowe
Editor: Paul Smaglik
Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Naturejobs Sales Director:
Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)
Andy Douglas (4975)
Frank Phelan (4944)
**Netherlands/ Italy/ Spain/
Portugal/ Belgium:**
Evelina Rubio Hakansson (4973)
Scandinavia: Silje Opstrup (4994)
France/ Switzerland:
Amelie Pequignot (4974)
Natureevents:
Paul Constant (4954)

Production Manager:

Billie Franklin
To send materials use London
address above.
Tel +44 (0) 20 7843 4814
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Office

Germany/ Austria:
Patrick Phelan, Odo Wulffen
Tel + 49 89 54 90 57 11/-2
Fax + 49 89 54 90 57 20
e-mail: p.phelan@nature.com
o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager:

Peter Bless

US Advertising Coordinator:

Linda Adam

Japan Head Office, Tokyo

MG Ichigaya Building (5F),
19-1 Haraikatamachi,
Shinjuku-ku,
Tokyo 162-0841
Tel +81 3 3267 8751
Fax +81 3 3267 8746
Asia-Pacific Sales Director:
Rinoko Asami
e-mail: rasami@naturejpn.com

naturejobs

Rewarding experience

The astute scientist recognizes that research is becoming increasingly multidisciplinary. This week, the London-based Royal Institution is honouring two scientists who are doing such work. Christopher Lowe and Martin Nowak will receive the Henry Dale Prize, funded by the Kohn Foundation, a charity that supports scientific/medical research.

It is perhaps appropriate that the Royal Institution is handing out the £10,000 (US\$16,700) prize, says Lowe, director of the Institute of Biotechnology at the University of Cambridge, because Michael Faraday, a pioneering multidisciplinary scientist, was strongly associated with the institute. "He did chemistry, physics and engineering," says Lowe. "We've lost that over the past 200 years, but now it's starting to come back again."

Lowe and Nowak, who directs Harvard University's programme for evolutionary dynamics, take different approaches to bringing multiple competencies into a research team. Lowe has assembled a team of about 100 scientists trained in different skills, with some intentional redundancy. The most important commonality, he says, is a willingness to communicate and learn about how colleagues' techniques could be applied to their problems.

Nowak recommends preparing for a cross-disciplinary career by switching track between a first degree and a PhD — from maths to biochemistry, for example. Nowak sees as positive the recent trend among US students of doing multiple major and minor subjects during their undergraduate years.

There is no one right way to pursue a multidisciplinary career (see *Nature* **425**, 542–543; 2003). But researchers who see that science is heading in that direction, and find the best way for them to follow it, should be rewarded with a successful career.

Paul Smaglik

Naturejobs editor



Contents

CAREERS AND RECRUITMENT

The rise of functional genomics p880

Bridging the technological gap p882

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

ANNOUNCEMENTS

EVENTS

In search of form and function

Rapidly changing technology and an abundance of DNA sequences are creating more job opportunities in functional genomics — particularly for scientists who have been trained outside traditional biology. Hannah Hoag investigates.

On the genetics timeline, the first complete sequence of a free-living organism, that of the bacterium *Haemophilus influenzae* in 1995, was finished 130 years after Gregor Mendel published the results of his famous pea-breeding experiment. Since that sequence, others have followed thick and fast, including those of the nematode *Caenorhabditis elegans*, the fruitfly, humans and now the dog. But each sequence alone cannot provide researchers with much information about the organism. For that, they have to go through the painstaking process of knocking out genes, one at a time, and observing how these deletions affect biological function.

But now that a number of genome sequences have been completed to reasonable accuracy, the field of functional genomics — finding out which proteins each gene produces and how they interact with one another — is becoming more robust. Governments are providing more funding for functional-genomics projects and training schemes, and international drug companies are eyeing the field eagerly.

Technology is also improving the science — and creating job opportunities for those who can master it. Advances in microarrays

and RNA interference, a technique for ‘silencing’ specific genes, is driving the field forwards (see ‘Interference science’, below).

“With the availability of genome sequences we have the ability to ask questions that we could not ask before,” says John Quackenbush, a specialist in genomic data analysis at The Institute for Genome Research (TIGR) in Rockville, Maryland. “They are a tremendous resource and people are now in the position to exploit them.”

Genomics research is moving from being data poor to becoming data rich, agrees Brigitta Tadmor, executive director of the Computational and Systems Biology Initiative at the Massachusetts Institute of Technology in Cambridge. But although functional genomics is one of the fastest-growing fields in biology, it takes more than a classical biology education to become part of the workforce — to capitalize on its potential benefits, it needs people with a blend of mathematics, biology and engineering.

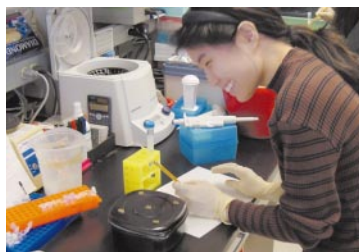


John Quackenbush: genome data present biologists with a new set of challenges.

TIGR

TIGR

Sowing the seeds: functional genomics is a major growth area for biologists.



Interference science

Andrew Fire and Craig Mello revolutionized gene-function analysis in 1998 when they injected double-stranded RNA into the nematode *Caenorhabditis elegans* and silenced the expression of genes that had known, but uncharacterized, counterparts in other organisms. The technique, now known as RNA interference (RNAi), is used to knock out large numbers of protein-coding genes within an organism, allowing researchers to do genome-wide genetic studies. The

technique is considered to be a direct way of assessing gene function. “RNAi is a revolution in biology and we are just starting to appreciate it,” says Norbert Perrimon, a geneticist at Harvard Medical School in Boston, Massachusetts.

Universities are setting up new facilities dedicated to RNAi screens (see ‘Building bridges’, page 882), and pharmaceutical firms such as Novartis, Bristol-Myers Squibb and Roche have added RNAi to their genomics technology toolbox. **H.H.**

THE DATA DILEMMA

If anyone is going to understand how genomes work, it will be a multidisciplinary, data-friendly team. Researchers agree that the hardest aspect of recruiting is finding people with backgrounds that let them both produce and interpret data. Biologists are essential because they understand the systems, but the technologies have advanced far beyond gene sequencing and simple gel electrophoresis. People with skills in biology and mathematics, or in biology and nanotechnology, might benefit most from the boom. Computational biologists, in particular, have been in demand.

At TIGR, bioinformaticians have been central to the institute’s success. “There is a lack of realization as to how challenging the data are. In physics the people who do the experiments are also writing the



Genomics on the move: across the lab, the changing face of technology is giving rise to a need for a broader range of skills than just classical biology.

software; in biology there is a disconnect,” says Quackenbush, who adds that biologists are slowly becoming more statistically savvy. Although biologists don’t have to be able to write software, they need to be able to communicate with the person who does. “I require my postdocs to take a programming class,” says Quackenbush. “It is important for biologists to gain the computational tools and techniques.”

But many other biologists still dismiss the importance of complex statistical analysis to their research. Roger Foxall, an executive vice-president of Genome British Columbia, a research organization based in Vancouver, Canada, that invests in genomics and proteomics research projects, says that he is astounded at the naivety of some biologists — nearly half of the grant applications he received recently didn’t mention bioinformatics. “The major thing is to view biology as an information science,” says Leroy Hood, president of the Institute for Systems Biology in Seattle, Washington.

SLIDING SCALE

Although biologists rely on biostatisticians to organize and analyse the data, they rely on physicists and chemists to improve the technology. A few years ago, ‘DNA microarrays’ were the words on everyone’s lips. These chips held thousands of genes and allowed scientists to see which ones were turned on or ‘expressed’ in different tissues or at different times. The technology is rapidly evolving — chips used to be difficult to make and often had to be bought commercially at great expense. Now, DNA microarray systems can print hundreds of slides in a single day.

In addition, the type of material held by the chips is expanding — microarrays can now capture antibody

interactions. And the next step is to develop chips that can measure interactions between proteins.

For example, a well-designed protein array could allow researchers to measure the effects of a drug binding to a targeted receptor. But the characteristics of proteins make them more difficult to work with as, unlike genes, they change shape. Also, whereas in humans there are tens of thousands of genes, there are hundreds of thousands of proteins.

These challenges mean that there are opportunities for people who can design new technologies. “You will need new chip surfaces and new types of readouts,” says Mike Taussig, chairman of the European Science Foundation’s Programme on Integrated Approaches for Functional Genomics. Linking the proteins to the chip surface is a problem best solved by chemists, whereas physicists will be best suited to designing the detection systems.

And now, genomics is now moving into the realm of nanobiology and nanoengineering. Biostatisticians may be popular today, but industry is already clamouring for engineers, physicists and chemists who can improve on the current technologies and develop new ones. Those enterprising individuals who can cross disciplines are encouraged to do so now, when their skills are most in demand. ■

Hannah Hoag is a science writer in Montreal, Canada.

Web links

Computational and Systems Biology Initiative

♦ csbi.mit.edu

Programme on Integrated Approaches in Functional Genomics

♦ www.functionalgenomics.org.uk

The Institute for Genome Research

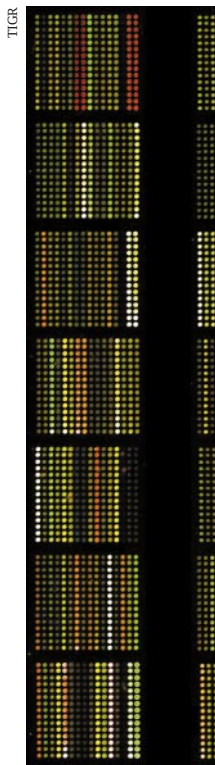
♦ www.tigr.org

National Human Genome Research Institute

♦ www.nhgri.nih.gov

Building bridges

The costs of functional genomics can be prohibitive, and job candidates often lack the skills most researchers desire, but many academic settings are creating training schemes and unique institutes to deal with these barriers. Hannah Hoag reports.



Microarray slides, here being prepared by a robot, are now allowing researchers to investigate protein interactions.

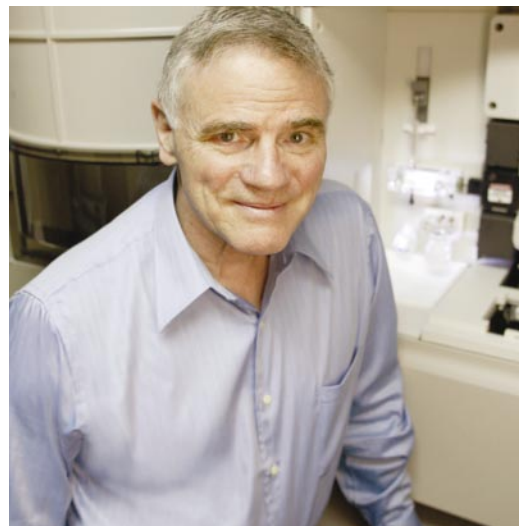


When high-throughput screening of proteins and genes made its debut, it was the darling of industry — but new academic facilities are now being set up to make the expensive technology involved available to researchers. As a result, new institutes are recruiting scientists to liaise between academia and industry; and industry, in turn, continues to seek graduates and postdocs — as long as they are fresh thinkers with multiple skills. In addition, centres are being set up to train students in the latest technologies.

A prime example focuses on RNA interference (RNAi), a useful technique for genetic analysis that works by 'silencing' specific genes. In May, Norbert Perrimon, a geneticist at Harvard Medical School in Boston, Massachusetts, launched the *Drosophila* RNAi Screening Center, the world's first RNAi laboratory. He says that he established the lab because of the prohibitive cost of carrying out RNAi screens. "If you compare RNAi to microarrays, there is a tenfold increase in complexity and cost," he says. The lab's primary funding source is a \$3-million grant from the National Institute of General Medical Sciences (NIGMS).

The centre is equipped with robots, plate readers and screening microscopes, and provides the methodology and tools to do high-throughput RNAi screens in tissue-cultured *Drosophila* (fruitfly) cells. "Researchers will come in for a week or two, then go back to their own labs with the genes that they have identified," Perrimon explains. Because the lab is still very young, Perrimon says that he won't be hiring many staff, but he does plan to build on his experience and open additional facilities in Canada and Germany. The key traits that he looks for in job candidates are an exceptional ability at informatics and a biology background.

Perrimon's RNAi lab is just one of the high-throughput screening facilities being set up in academia across the United States. The directors of these new facilities hope to accelerate the pace of functional-genomics research. Anindya Dutta, a professor of biochemistry and molecular genetics, recently moved from Brigham and Women's Hospital in Boston, Massachusetts, to the University of Virginia in Charlottesville to set up an RNAi screening facility to study cell proliferation in human cells. Dutta says that he plans to hire four young



Leroy Hood wants to see strong cross-disciplinary efforts in genomics.

scientists, either graduate students or postdocs, but he is worried about whether he can attract bright people with informatics experience. "Software people expect higher salaries, but it doesn't always fit into a National Institutes of Health grant," he says.

EASTERN OPPORTUNITIES

Another research institute that is moving into functional genomics is the Genome Institute of Singapore (GIS). Because of the government's commitment to designing new genomics technologies and funding genome-based research, the GIS has become a flagship programme in Singapore's biomedical-sciences initiative, says Lance Miller, a senior group leader in microarray and expression genomics at the institute.

Miller says that there are numerous job opportunities at the GIS and elsewhere in Singapore — so much so that demand is outstripping supply, with institutes often competing with each other for the same candidate. Miller says that the most desirable candidates are biologists who can embrace mathematics. "It helps to have certain computing, informatics or statistical skills," he says. "They can carry smart biologists a long way in the field of functional genomics."

Mathematics and statistics are so integral to functional genomics that a third of the Institute for Systems Biology (ISB) in Seattle, Washington, is

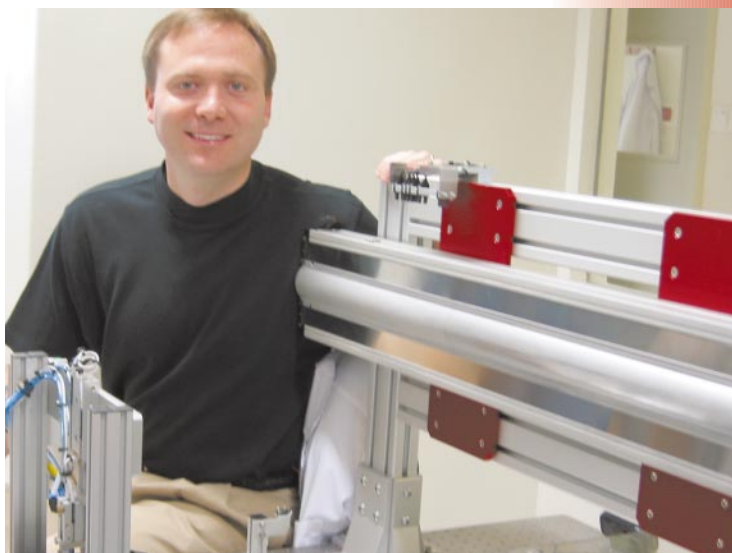
devoted to their use, says Leroy Hood, president and founder of the institute. He formed the ISB in 2000 because he believed that it would be difficult to set up the required infrastructure in a traditional academic environment (see *Nature* **407**, 828–829; 2000). Biologists needed to interact with engineers and statisticians to modify existing technologies and to develop new ones. “We needed cross-disciplinary scientists who were willing to understand the biology and to tackle the deep problems,” he says.

One of the institute’s goals is to develop new genomics, proteomics and computational tools to analyse the interactions of proteins within cells. The ultimate goal is to develop better diagnostic tools and drugs to circumvent defects in these protein networks. Hood says that he expects to hire three or four faculty members with backgrounds covering physics, chemistry, mathematics and biology over the next couple of years. Senior scientists, postdocs and graduates are always in demand, and the institute hires several undergraduates on temporary contracts over the summer, as long as they have strong engineering or mathematics backgrounds.

ADDRESSING THE DEFICIT

The paucity of adequately trained graduates has led some universities to set up academic departments devoted to functional biology. At the Computational and Systems Biology Initiative (CSBi) at the Massachusetts Institute of Technology in Cambridge, faculty members and students started a grass-roots campaign to make sure that students would graduate with job offers in their hands. The institute will establish six core competencies including molecular genetics and genomics, proteomics and structural biology, imaging and imaging informatics, and microsystems.

The centre has attracted plenty of attention for its efforts. Drug firm Merck has donated \$2 million to help train graduates ready for the biotechnology industry, and Dell and IBM have offered computer clusters and technical-support specialists. The centre



Lance Miller says that biologists will find a wealth of job opportunities in Singapore.

has also secured a \$16-million grant from the NIGMS to develop a programme in microsystems. “We hope to announce the first PhD programme in January and have the first class in autumn 2004,” says Brigitta Tadmor, executive director of the initiative. Six faculty members are being recruited to develop the research infrastructure, train staff and students, and to create industrial partnerships.

The CSBi hopes to meet the needs of pharmaceutical companies such as Roche, headquartered in Basel, Switzerland. Many employees come to Roche during graduate school, providing them with an early start towards a career in industry. Klaus Lindpaintner, vice-president of research at Roche Genetics Europe, says that the company needs people with technical skills, such as RNAi and mass spectrometry, but that statistics know-how is essential. “We need people with statistics training in three subdisciplines,” he says. “There are those who are architects and engineers capable of putting together the data-collection warehousing systems, and those who are geared towards using those databases and generating algorithms. The third breed are those who really understand how to analyse the data, the advanced statistical geneticists. They are in short supply.” Roche has about 2,000 preclinical research staff worldwide, and usually sees an annual turnover of 10%.

Traditionally, university graduates with a speciality in a particular area or technique have been able to find jobs that match their skills and experience. But as the data begin to accumulate, mathematics is becoming more integral to functional genomics. Researchers warn that experts — no matter how skilled — will lose out to those who understand the overall picture. With both industry and academia on the lookout for graduates with backgrounds in biology and mathematics or engineering, enterprising biologists should develop their mathematics skills. And statisticians looking for a growing and evolving problem might try their hand at functional genomics. “Industry is clearly looking for people who are educated in this engineering–biology interface,” says Tadmor. “Not just someone who learned to do some data processing, but someone who can think in both disciplines.”

Hannah Hoag is a science writer in Montreal, Canada.

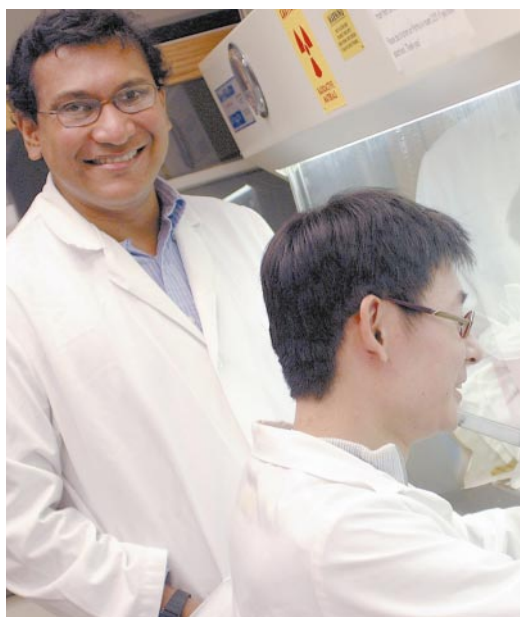
Web links

Roche

♦ www.roche.com

♦ *Drosophila* RNAi Screening Center

♦ flyrnai.org



Anindya Dutta (left) is using RNA interference to study cell proliferation.